

B 1573

M6

#8

922

3 1573
M
UNIVERSITY OF ILLINOIS BULLETIN

ISSUED WEEKLY

Vol. XIX

JANUARY 30, 1922

No. 22

[Entered as second-class matter December 11, 1912, at the post office at Urbana, Illinois, under the Act of August 24, 1912. Acceptance for mailing at the special rate of postage provided for in section 1103, Act of October 3, 1917, authorized July 31, 1918.]

BULLETIN NO. 8

BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION

A CRITICAL STUDY OF CERTAIN
SILENT READING TESTS

By

WALTER S. MONROE, DIRECTOR



PRICE 50 CENTS

—
PUBLISHED BY THE UNIVERSITY OF ILLINOIS
URBANA

Monograph

BULLETIN NO. 8
BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION

A CRITICAL STUDY OF CERTAIN SILENT READING TESTS

By
WALTER S. MONROE, DIRECTOR



PUBLISHED BY THE UNIVERSITY OF ILLINOIS
URBANA

LB 1573
M6

LIBRARY OF CONGRESS
RECEIVED
OCT 18 1922
DOCUMENT

TABLE OF CONTENTS

	PAGE
PREFACE	4
THE MEASUREMENT OF SILENT READING ABILITY.....	5
THE PROBLEM.....	5
THE DATA COLLECTED.....	6
THE PERFORMANCES REQUIRED OF A PUPIL.....	9
DESCRIPTION OF PUPILS' PERFORMANCES.....	13
SCORING REPRODUCTIONS.....	15
THE IDEA-COUNTING METHOD.....	16
BROWN'S METHOD OF IDEA COUNTING.....	17
THE WORD-COUNTING METHOD.....	19
SUBJECTIVITY OF DESCRIBING REPRODUCTIONS.....	20
CONSTANT ERRORS AND VARIABLE ERRORS.....	20
SUMMARY FOR DESCRIBING REPRODUCTIONS.....	25
SCORING ANSWERS TO QUESTIONS.....	25
DESCRIBING THE QUALITY OF COMPOSITIONS.....	25
TIME REQUIRED FOR SCORING TEST PAPERS.....	25
AVERAGE SCORES AND STANDARD DEVIATIONS.....	26
EQUIVALENCE OF DUPLICATE FORMS.....	29
RELATION OF VOCABULARY TO DIFFICULTY.....	30
FORMATION OF COMPOSITE SCORES.....	31
RELIABILITY	32
METHODS OF DETERMINING RELIABILITY.....	32
PROBABLE ERROR OF t DUE TO SAMPLING.....	33
RELIABILITY OF TESTS STUDIED.....	34
DISCRIMINATION	37
COMPARISON WITH TEACHERS' RATINGS.....	39
CORRELATION OF COMPREHENSION WITH MEMORY	39
CORRECTED COEFFICIENTS OF CORRELATION.....	41
CORRELATION OF COMPREHENSION WITH VOCABULARY.....	42
CORRELATION OF CANCELLATION SCORES WITH MEASURES OF RATE OF READING.....	44
CORRELATION OF COMPREHENSION WITH WRITTEN COMPOSITION	44
INTER-CORRELATION BETWEEN TESTS.....	46
CORRELATION OF SINGLE TESTS WITH COMPOSITES.....	50
SUMMARY OF CONCLUSIONS.....	51
CORRELATION WITH COMPOSITES.....	52

PREFACE

In the field of silent reading, as well as in the fields of other school subjects, the number of available educational tests has been increased so that one desiring to use a test is confronted with the necessity of making a choice. If such a choice is to be made intelligently it is necessary to have at hand experimental data with reference to the reliability and validity of the tests considered. The study which is reported in this monograph was undertaken for the purpose of securing such data with reference to certain silent reading tests. The report is presented in hopes that users of silent reading tests will find the information that it contains helpful in making an intelligent selection of educational tests in this field. The monograph will doubtless also be of interest to students in the field of educational measurements.

WALTER S. MONROE,
Director, Bureau of Educational Research.

A CRITICAL STUDY OF CERTAIN SILENT READING TESTS

The measurement of silent reading ability. The scores yielded by silent reading tests may fail to be true measures of silent reading ability for two reasons. *First*, the scores may not be reliable or accurate. A score is lacking in reliability when two applications of a test or of duplicate forms of it do not yield approximately the same score when administered to the same pupils, as far as possible, under the same conditions. Included in this is any lack of objectivity in the scoring of the test. *Second*, the performance which a pupil gives on a silent reading test may depend upon other factors in such a way that it is an index of these factors rather than of silent reading ability. For example, when a pupil answers questions from memory his answers may be influenced to such an extent by his ability to remember that his performance is not a truthful index of his ability to read silently.

Two aspects of the activity of silent reading may be recognized. *First*, the reading mechanism consists of perception, eye-movement habits, etc. The rate of silent reading is largely dependent upon this mechanism and hence any measure of rate is an index or symptom of the quality of the mechanism. *Second*, the thought-getting or comprehension aspect of silent reading involves the higher mental processes. The quality of this is indicated by the comprehension scores. Comprehension is not entirely independent of the mechanism of silent reading, but, if sufficient time is allowed, pupils who possess poor reading mechanism may stand high in thought-getting.

The problem. The problem of this study is to ascertain the reliability and, so far as possible, the function and validity of certain silent reading tests. These tests, as will be shown later, differ in the performances which are required of the pupils. They also differ in other respects. Their titles suggest that all of the silent reading tests included in this study are designed to measure silent reading ability. The fact that they differ widely in certain respects suggests the possibility that no two of them measure the same type of reading ability, or at least that they do this with different degrees of validity. The study has been restricted to tests which yield some measure of the rate of reading as well as a measure of comprehension in order that the measurement of both phases of silent reading

activity might be studied. With one exception, the tests which have been used have duplicate forms. In addition to the silent reading tests, certain other tests were given to the same pupils, because it was thought that the scores yielded by them might assist in the analysis and interpretation of the scores yielded by the silent reading tests.

The data collected. Through the courtesy of Superintendent W. W. Earnest and certain teachers of the Champaign Public Schools, the tests chosen for this study were given in the spring of 1920 to a number of pupils in the fourth and seventh grades. All of the tests were administered by Miss Dora Keen, at that time a research assistant in the Bureau of Educational Research. Care was exercised to secure as nearly uniform testing conditions as can be obtained in the ordinary schoolroom. The lapse of time between the giving of the different forms of the same test was made as nearly equal as possible for the different groups. Only in rare instances were tests given after recess in the afternoon or during the afternoon session on Friday. The tests were given to all pupils in four rooms in both the fourth and seventh grades. The total number of pupils tested in each grade was approximately 140. The study is, however, based upon the records of only those pupils who took all of the tests. The number of complete records in the fourth grade is 80 and in the seventh grade, 91.

The following tests were given in the fourth grade:

1. The Curtis Silent Reading Test No. 2¹, Form 1, "The Kitten Who Played May Queen," and Form 3, "The Kitten Who Caught a Fish."

2. Brown's Silent Reading Test², Form 1, "The Long Slide," and Form 2, "A Morning Adventure."

3. Monroe's Standardized Silent Reading Test I³, Forms 1, 2, and 3.

¹Curtis Silent Reading Test No. 2. Forty-sixth Annual Report. Kansas City, Missouri: Board of Education, 1917. pp. 79-85.

²Brown, H. A. "The Measurement of Ability to Read." A Manual of Directions Concerning Giving and Scoring of Reading Tests, Statistical Treatment of the Data and Diagnosis of School Class and Individual Needs. Concord: New Hampshire Department of Public Instruction (in cooperation with the General Education Board). Bureau of Research Bulletin No. 1, Second Edition, 1916. pp. 57.

³Monroe, W. S. "Monroe's Standardized Silent Reading Tests." *Journal of Educational Psychology*, 9:303-12, June, 1918.

4. Fordyce's Scale for Measuring Achievement⁴ in reading Test No. 1, "Narcissus."

5. Experimental Reproduction Test I, Form 1, based on pages 84 and 85 of the supplementary reader, "The Strike at Shane's"⁵, and Form 2 based on pages 6 and 7 of the same publication. The passage for Form 1 contains 370 words and that for Form 2, 395 words. In administering these tests the pupils read from the supplementary reader. The exact place of beginning had been marked in each copy. Also the end of the passage to be read was indicated.

6. Cross-Out Silent Reading Test I, Form 1 and Form 2. This is an experimental silent reading test. In a passage of rather simple reading material, words were substituted, which did not agree with the meaning of the preceding words in the sentence. A pupil is asked to cross out the words which do not fit. With the exception of the substituted words, the selection is a connected story.

7. Vocabulary Test. The words of this test are those used by Terman and Childs. The form of the test is that proposed by Whipple⁶.

8. Cancellation Test, "a-t" and "e-r"⁷.

9. Memory, "How Mr. Lincoln Helped the Pig."⁸

The following tests were given in the seventh grade:

1. Starch's Silent Reading Test No. 6 and Test No. 7.⁹

2. Monroe's Standardized Silent Reading Test II, Forms 1, 2, and 3.

3. Fordyce's Scale for Measuring Achievement in Reading, Test No. 2, "Spirit of Spring."

⁴Fordyce, Charles. "A Scale for Measuring the Achievements in Reading." The University Publishing Company, Lincoln, Nebraska, and Chicago. 1916.

⁵"The Strike at Shane's." (Gold Mine Series, No. 2.) Boston: American Humane Education Society, 1908. pp. 91.

(A supplementary reader for the fourth grade which has as its lesson kindness to domestic animals.)

⁶Whipple, G. M. Manual of Mental and Physical Tests, Complex Processes Chapter 12. Baltimore: Warwick and York, 1914.

⁷This test is described by Whipple in the Manual of Mental and Physical Tests, Simpler Processes, p. 311.

⁸Whipple, G. M. Manual of Mental and Physical Tests, Simpler Processes, Pages 207-10.

⁹Starch, Daniel. The Measurement of Efficiency in Reading. Journal of Educational Psychology, 4:1-24, 1915. These tests were used as duplicate forms.

4. Experimental Reproduction Test II, Form 1, based on pages 6 and 7 of the supplementary reader, "Old English Heroes,"¹⁰ and Form 2, based upon pages 8 and 9 of the same publication. The passage for Form 1 contains 662 words, and that for Form 2, 611 words.

5. Cross-Out Silent Reading Test II, Form 1 and Form 2. This test is similar to the Cross-Out Silent Reading Test used in the fourth grade but is based upon more difficult material.

6. Pressey Silent Reading Test for Grades VI, VII, and VIII, Form 1 and Form 2. This is an experimental test.

7. Vocabulary Test. This is the same test as that used in the fourth grade.

8. Cancellation Test, "a-t" and "e-r." This is also the same as that used in the fourth grade.

9. Memory Test, "Marble Statue."¹¹

10. Composition Test. The Willing Composition Scale¹² and the directions which accompany it were used.

In addition to the above tests a rating for ability in silent reading was secured from the teachers. To guide them in making this rating, the teachers were given the following directions:

Think of all the fourth (seventh) grade pupils with whose silent reading ability you have ever become acquainted from the best to the poorest. Compare each child in your present class with this distribution of pupils. Give a pupil a rating of 5 if he has *very superior* ability in silent reading equalled only by about seven out of every hundred, or 7 percent of fourth (seventh) grade pupils. Give him a rating of 4 if he has *superior* ability or ability above the average, yet is excelled by the *very superior* group. About 24 out of every hundred, or 24 percent of fourth (seventh) grade pupils, will fall in the *superior* group. Give him a rating of 3 if he possesses average ability, i. e., ability which lies somewhere close to the middle of the difference between the very best pupil and the very poorest. About 38 out of every hundred, or 38 percent of fourth (seventh) grade pupils, will fall in this average group. If the pupil is below the average in ability to read and yet

¹⁰Bush, Bertha E. Old English Heroes. (Instructor Literature Series—No. 116.) Danville, N. Y., and Chicago: F. A. Owen Publishing Co., and Hall and McCreary, 1909. Pp. 31.

This is a supplementary reader suitable for the upper elementary grades. It contains brief sketches of the lives of Alfred the Great, Richard the Lion-Hearted, and the Black Prince.

¹¹Whipple, G. M. Manual of Mental and Physical Tests, Simpler Processes, Pages 107-10.

¹²Willing, M. H. Measurement of Written Composition in Grades IV to VIII, English Journal, 7:193-202, March, 1918.

does not equal the poorest you have ever known give him a rating of 2. This group is called *inferior* and will contain about 24 out of every hundred, or 24 percent of fourth (seventh) grade pupils. Give the pupil a rating of 1 if he is *very inferior* in ability to read so that he is as poor or very nearly as poor as the poorest pupil you have ever known. About 7 out of every hundred, or 7 percent of fourth (seventh) grade pupils, will fall in this very inferior group.

The above directions do not mean that you will necessarily be obliged to give 7 percent of your class a rating of 5; 24 percent, a rating of 4; 38 percent, a rating of 3; 24 percent, a rating of 2; and 7 percent, a rating of 1. They do mean, however, that a large number of pupils, a number running up into the hundreds, can be divided in exactly this manner, i. e., 7 percent, very superior; 24 percent, superior; 38 percent, average; 24 percent, inferior; and 7 percent, very inferior. You are to think of all the pupils you have ever known from the best to the poorest and by comparison give each pupil in your present class the rating he would receive if he were included with all the pupils you have known and the entire number should be rated in the above manner.

The performances required of a pupil. All of the silent reading tests in the above list are designed to measure the ability to read silently. However, they require a variety of performances from the pupil. In the Courtis Silent Reading Test No. 2, the pupil is required to read a continuous selection for three minutes. At the end of this time he turns to another section of the test and answers questions based upon the selection he has just read. The questions are to be answered by either "yes" or "no." The selection read is repeated in connection with the questions so that the pupil may refer to it in case he does not remember the answer to any question. The Brown Silent Reading Test and the Starch Silent Reading Tests require the pupil to read a selection and then reproduce what he can remember. Starch allows thirty seconds reading time, while Brown allows one minute. The Monroe Standardized Silent Reading Tests consist of a series of exercises. Each exercise consists of one paragraph and a question based on it. Most of the answers are to be given by drawing a line under a word. Five minutes are allowed for the test. The Fordyce Scale for Measuring Achievement in Silent Reading¹³ requires the pupil to read a selection and then answer from memory questions based on it. The selection for Test 1 contains 300 words. The time allowance is 125 seconds. The selection for Test 2 contains 512 words with a time allowance of 140 seconds. The time allowed for the reading is intended to be such that 50 percent of the pupils will finish before time is called. The directions which accompany the Fordyce Scale for Meas-

¹³This test has only one form. Test 1 was given in the fourth grade and Test 2 in the seventh grade.

uring Achievement in Silent Reading are stated in general terms. For this reason it was necessary to formulate the exact explanation to be given to the pupils. The following was used:

Do not turn over your paper until I tell you to begin. These papers have a story on them. You are to read the story at your ordinary rate of reading, carefully enough so that you will be able to reproduce the leading thoughts. When I say "mark," draw a line around the word at which you are looking at that time. If you have not finished go right on reading until you come to the end of the story. Then immediately turn your paper face down and sit quietly until all have finished. You are to read the story once and once only, and just as soon as you have finished, turn your paper down. Is there any one who does not understand exactly what to do? All right! Begin!

In the Experimental Reproduction Test the following directions were used:

Do not open your books until I tell you to begin. Write your name and school on the card.¹⁴

This is a test to find out how rapidly and how well you can read. Read carefully; for you will be asked to write out what you have read. Put your finger in the book this way (illustrating). When I say "begin" open your books and begin to read at the first blue mark here (illustrating). When I say, "mark," draw a line around the word at which you are looking, (illustrate), then go right on reading until you come to the last blue mark. Then close your book and sit quietly until all have finished. Read over only once. Do not forget to draw a line around the word where you are reading when I say, "mark." Is there anyone who does not understand just what he is to do? All right! Begin!

The time allowance was thirty seconds. After they had completed the reading, the pupils were asked to write, in as nearly the same words as possible, all that they had read. This reproduction completed, they were asked to answer a list of questions based upon the selection read. They were not given an opportunity to consult the reproduction nor to add to it after answering the questions.

The nature of the Cross-Out Silent Reading Test is illustrated by the directions given to the seventh grade pupils:

Below you will find a paragraph of a story. Certain words in this paragraph do not belong there, that is, they do not make sense and do not agree with what has gone before. Read this paragraph carefully and draw a line through all the words which do not belong there. Do not write anything. Do nothing except cross out the words which do not make sense with what has gone before. Is there anyone who does not understand what he is to do? Remember to cross out only the words which do not agree with what has gone before. All right! Go ahead!

¹⁴A 3x5 card was fastened to the copy of the supplementary reader which was given to each pupil. Before the books were distributed to another class the rate scores were recorded on the cards and a new card attached.

"It happened in our country long ago, in those old days when only a few white people lived here and everything was rough and civilized. Strong men were at work among the hills, cutting down the brooks and planting corn in the new fields, and towns were springing up all along the walls, but still there were many miles of forest where Indians hunted and bears and wolves had their palaces."

In this paragraph the words to be crossed out are "civilized", "brooks", "walls", and "palaces". These answers were read to the pupils after they had marked the paragraph. In case any failed to understand the nature of the exercise it was explained to them. They were then directed as follows:

In the following pages you will find part of a story. It is not a fairy story. In this story, as in the paragraph above, there are words which do not agree with the meaning of what has gone before. Cross them out just as you did in the above paragraph. Be sure to cross out all the words which do not belong, but cross out only those words; for if you cross out any word which should not be crossed out it will be counted as a mistake. You will be allowed four minutes to work. Many of you will be unable to finish during this time. It is more important, however, to do your work correctly than to cover a great deal of ground. Do all three pages.

When I say "begin" turn the page and start to work. If anyone finishes before the time is up, close your paper and sit quietly. Is there anyone who does not understand just what he is to do? All right! Begin!

The directions to the fourth grade pupils differed from the above in only two respects. Two additional illustrative paragraphs were used and the time allowance was three minutes instead of four.

The nature of the Pressey Silent Reading Test for Grades six to eight may be illustrated by the directions:

Look at the first example given just below:

1. February is the longest month in the year. The above statement is not true; but there is only one word that makes the sentence untrue. This one word is the word "longest"; if "longest" were changed to "shortest", the sentence would then read, "February is the shortest month in the year", which is true. "Longest" is wrong; so take your pencils and cross it out. Draw a line through it because it is wrong.

Look at the second example just below:

2. The day dawned bright and dreary; the clear morning light streamed in through the windows and filled the room with its cheery brightness.

In this paragraph, also, there is one, *and only one*, word that is wrong, the meaning of which does not fit in with the meaning of the rest of the paragraph. The word is "dreary". Cross it out.

Two additional illustrative exercises were given and the pupil directed as follows:

And now—everyone attention! In each of the paragraphs on the other side of the page, there is one, *and only one*, word that is wrong, which makes the paragraph untrue, or whose meaning does not fit in with the meaning of the rest of

the paragraph. Cross that word out. And remember, there is only *one* word in each paragraph that is wrong. Be sure to take the paragraphs in order. Never skip a paragraph without attempting it. Read rapidly and accurately. You will be given 10 minutes in which to work. Ask no questions.

Now, turn over the page, and all start!

In the vocabulary tests the following directions, which are printed on the test papers, were read to the pupils:

Below are 100 words which are designed to measure the size of your vocabulary. Consider each one carefully, and place *before* it one of these four marks:

(1) the mark "D" if you could define it as exactly as words are ordinarily defined in the dictionary.

(2) the mark "E" if you could explain it well enough to give some idea of its meaning to one who is not familiar with it, though you could not give an exact definition that would satisfy an expert.

(3) the mark "F" if the word is merely roughly familiar, so that you have only an indefinite idea of its meaning and could not use it intelligently.

(4) the mark "N" if the word is entirely new and unknown to you.

When you have finished, count the marks and fill out these blanks, making sure that the numbers add to one hundred.

In the fourth grade these directions were modified somewhat in order to make certain that the pupils would understand them. Fifteen minutes were allowed for the test in both grades.

The Cancellation Tests consist of a page of Spanish text. For the "a—t" test the following directions were given to the pupils:

On this paper you will find a large number of words from a foreign language. Draw a line through each of these words which contain both an "a" and a "t."

If the word has an "a" but not a "t" in it do not cross out the word. If it has a "t" but not an "a" do not cross it out. Be sure to draw a line through all words which contain both an "a" and a "t," but only through these words; for if you cross out a word which does not have both an "a" and a "t" in it, it will count as a mistake. When I say "begin" turn over your paper and begin work. You will be allowed two minutes to work. Your score will depend on the number of words you cross out correctly.

In addition to this explanation of the test, four non-consecutive words were selected from the text and written on the blackboard in order to illustrate the kind of words to be crossed out. The explanation for the "e—r" test is identical with the above except that "e" and "r" are used in the place of "a" and "t."

In the Memory Tests the pupil was directed as follows:

This is to be a test to see how well you remember what you hear. I am going to read a little story, and I want every one to pay close attention; for as soon as I have finished I want you to write down, in as nearly the same words as possible, what I have just read to you. Listen carefully, and as soon as I stop reading write down all that I have just read. Your score will depend on how nearly you remember what has been read to you. Do not begin to write until I have finished

reading. Is there anyone who does not understand just exactly what he is to do? All right! Attention!

In the composition test the following topics were written on the blackboard. Then the directions given below were read to the pupils:

AN EXCITING EXPERIENCE.

A storm.	An unexpected meeting.
An accident.	In the woods.
An errand at night.	In the mountains.
A wonderful story.	On the ice.
A runaway.	On the water.

I want you to write me a story. It is to be a story about some exciting experiences that you have had, or about something very interesting that has happened to you. If nothing of the sort has ever happened to you, then tell me of an exciting experience someone whom you know has had. You may even make up a story of this kind, if you have to, though I believe you will do better, on the whole, with a real one. I am going to give you about twenty minutes in which to write. You are to write on both sides of the paper, to do all the work yourselves, and to ask no questions at all after you begin. You may make whatever corrections you wish between the lines. There will be no time to rewrite your story.

I have written the general subject on the blackboard, together with some suggestions. You do not have to write on any of these topics unless you want to; they are merely to help out in case you cannot think of an exciting experience yourself. Is there anyone who does not understand just what he is to do? All right! Begin!

Twenty minutes were allowed for the actual writing. Then the pupils were directed as follows:

You are to have four or five minutes in which to finish your stories, make corrections, and count the number of words written. Write this number at the end of your story.

Description of pupils' performances. In order to eliminate or reduce accidental errors and subjective errors to a minimum, all test papers were scored independently by two persons working under careful supervision. In the case of those scores for which the subjective factor was negligible, any differences between the two scores were reconciled by a third person.¹⁵ When a subjective error was involved the average of the two scores was taken unless the difference between them exceeded a fixed maximum. In this case the paper was scored by a third person in an attempt to reconcile the two scores.

The description of a pupil's rate of reading is objective. Hence only accidental errors are involved. The rate was expressed in terms of words per minute. The scoring of comprehension in the

¹⁵This third person was the same for all tests, and also was the one who supervised the scoring.

following tests was also highly objective: Monroe's Standardized Silent Reading Tests, Curtis' Silent Reading Test No. 2, Cross-Out Silent Reading Tests, Pressey's Silent Reading Test, and Cancellation Test.

Monroe's Standardized Silent Reading Tests were scored for comprehension according to the usual directions with a few slight changes with respect to the answers which were considered correct. The pupil's comprehension score is the sum of the comprehension values of the exercises which he does correctly.

The directions which accompany the Curtis Silent Reading Tests No. 2, provide for two measures of comprehension, the index of comprehension and the number of questions answered. The index of comprehension is found by subtracting the number of wrong answers from the number of right answers and dividing the difference by the number of right answers. In addition to these two scores the number of right answers was recorded.

Two methods of scoring the Cross-Out Silent Reading Tests for comprehension were used. It was found that pupils made two types of errors. Some crossed out words which should not have been crossed out, and words which should have been crossed out were not marked. One description was obtained by taking the difference between the number of words correctly marked and the number of words wrongly marked. (This included only the first type of error.) This score is indicated by the symbols $c - w$. In the second score, the number of inconsistent words, which the pupil failed to mark in the part of the test read, was recognized.

The score was obtained by evaluating the following fraction, $\frac{c - w}{c + o}$. In this fraction c and w have the same meaning as above and o stands for the number of words omitted.¹⁶

In the Pressey Silent Reading Test a pupil's comprehension score is the number of exercises which he does correctly within the time allowed. In order to have an exercise counted as right the correct word must be crossed out and no other word in the paragraph marked.

The Vocabulary Test was scored according to standard directions.¹⁷ Each "D" and "E" was regarded as indicating one point and each "F" as indicating a half-point. (See page 12.) The total

¹⁶Whipple, G. M. Manual of Mental and Physical Tests. Simpler Processes, p. 313.

¹⁷Whipple, G. M. Manual of Mental and Physical Tests, Part II, Complex Processes, p. 310-11.

number of points represents a vocabulary-index. This index, taken as a percent and multiplied by 18,000, affords a measure of the size of the pupil's total vocabulary.

In the cancellation tests the score was obtained by converting rate and accuracy into a single index of efficiency (E).¹⁸ This index was obtained by the following formulae:

$$A = \frac{c-w}{c+o} \quad E = e A$$

Here A = the index of accuracy.

E = the index of net efficiency.

e = the number of words examined.

o = the number of words erroneously omitted.

c = the number of words crossed.

w = the number of words wrongly crossed.

After computing the index of accuracy the score in terms of the index of efficiency was obtained.

The scoring of answers to questions obtained from Fordyce's Scale for Measurement of Achievement in Silent Reading and from the Experimental Reproduction Tests is less objective than the scoring of the tests just described. Fordyce gives a list of correct answers. This, together with the nature of the questions, makes the scoring of his test highly objective for its type. In the course of scoring the answers to the questions of the Experimental Reproduction Tests, lists of correct answers were compiled and all scoring was done in accordance with them. The acceptable answers were chosen with care from the complete array of all answers given in each of the tests. Any word or group of words judged to give correctly the total idea called for by the question was counted as correct.

Scoring Reproductions. The reproductions obtained from Brown's Silent Reading Test, Starch's Silent Reading Tests, the Experimental Reproduction Tests, and the Memory Tests were scored by both the "idea-counting method" and the "word-counting method." In addition, Brown's tests were scored according to the directions which he gives. The description of a reproduction is not highly objective. Pupils differ widely with respect to vocabulary and to sentence structure. In addition to incorrect statements, reproductions contain superfluous statements and repetitions. The order of ideas is frequently transposed so that their significance is modified. Ideas contained in the passage read are expressed with

¹⁸Whipple, G. M. Manual of Mental and Physical Tests, Part I. Simple Processes, pp. 312-13.

various degrees of completeness. These characteristics of reproductions create many opportunities for differences of opinion in their description.

1. The idea-counting method. The first step in using this method is to divide the selection read into ideas. In making this division one may adopt a relatively small unit, which is essentially a word or phrase, or a large unit, which approximates a sentence. After experimenting with these two plans of division the former was chosen. A portion of Brown's Silent Reading Test, "The Long Slide," with the divisions indicated, is reproduced below:

THE LONG SLIDE

The boys / and girls / who live / in a certain part / of a small / town / in the country / several miles / from any village / attend / school / in a little / red / schoolhouse / known as / the Long Hill / school. /

It has / this name / because / it is situated / on the top / of a very long / steep / hill. / Ever since anyone / can remember, / the scholars / of the Long Hill / school / have always had / time / to slide / down the hill / just once / at recess / in winter / and get back / to the school house / before the bell / rings / to call them back again / into school. / They can go down / very rapidly, / but it takes / a long time / to walk back. /

Last Monday / morning / Frank Lane / appeared / at school / with a fine / new / sled. / It was a double-runner / which his uncle, / who owns / a carriage factory / in the city, / had given him. / He named / his new / sled / the Simoon / and almost had / a fight / with Tom Smith, / who said / it was foolish / to put / such a name / on a sled, / but he kept on / calling it / the Simoon. /

At recess / that day / Frank / invited / the whole / school / to go / for a coast / and the twelve / boys / and girls / got onto / the sled / and away they went / down the steep hill. / When recess was over / Miss Black, / the teacher, / rang the bell / but not a scholar / appeared. / Thinking that / the children / had stopped / to play / on the way back / from their slide, / Miss Black / went / to the door / and looked / down the hill / and rang / the bell / again. / But not a scholar / was in sight. / Then she was greatly astonished / and began / to be very angry, / for nothing / like this / had ever happened / in all of her twenty-eight / years / as a teacher. / She waited / and waited / but still / no scholars / appeared. / She stopped / every team / that came / up the hill, / but no one / had seen / anything of them. /

She stayed / at the schoolhouse / and wondered / what had become of / her children / until it was time / to let out / school / and then / she went / over to John Reed's / who lives / nearest to the school house / and whose son / and daughter / were among the missing / scholars. / Mr. Reed / was greatly frightened / at what Miss Black / told him / about the disappearance / of her school / and immediately / hitched up / his horse / to go in search / of the lost / children. / Just / as he was driving / out of the dooryard / the scholars / appeared / far down the hill. / It was almost / dark / before / they got back / to the schoolhouse. /

The pupil's score is the number of ideas which he reproduces correctly. Thus, the scorer must determine what ideas, occurring in the passage read, appear in the pupil's reproduction. Two rules were adopted.

1. Misplaced clauses and phrases, that is, clauses and phrases which are tacked on to the wrong part of a sentence, are to be counted as incorrect.

2. Correct ideas found in a statement, which, as a whole, is directly contrary to the meaning of the text read, are to be counted

as correct. The following example may be cited: *John Shane was not cruel*. Here, both the ideas, *John Shane* and *cruel*, are held to be correct, while *was not* is incorrect. In practically all cases coming under this rule the incorrectness of the statement was caused by the use of a wrong verb or a wrong adverbial modifier, as in this illustration.

The scorers were urged to keep in mind the general rule that they were to match up identical ideas in the passage read and in the pupil's reproduction, even though sometimes the ideas were not expressed in the same language. In order to secure independent scorings, each selection, with the divisions into ideas indicated as shown above, was mimeographed. The scorer indicated on this mimeographed copy the ideas which in his judgment the pupil had reproduced. In this way no record of the scoring was made on the pupil's test paper, and complete independence of scoring was secured.

In putting together the results from two independent scorings, when the difference in the number of ideas was six or less, the average was taken. In the case of a difference of more than six the third person went over both papers to change too lenient or too severe scoring. These changes were made until the difference was reduced to six or less. Then the average was taken.

Brown's method of idea-counting. Brown has given directions for describing the reproductions written by pupils in terms of "quantity of reproduction" and "quality of reproduction." As a basis for his method of scoring, the selection is divided into sections each of which he considers to represent a unit of thought. A portion of "The Long Slide" is reproduced to show his plan of division:

THE LONG SLIDE

The boys and girls who live in a certain part of a small town in the country several miles away from any village attend school(1) in a little red schoolhouse known as the Long Hill School.(2)

It has this name because it is situated on the top of a very long, steep hill.(3) Ever since anyone can remember, the scholars of the Long Hill school have always had time to slide down the hill just once at recess in winter and get back to the schoolhouse before the bell rings to call them back again into school. They can go down very rapidly, but it takes a long time to walk back.(4)

Last Monday morning Frank Lane appeared at school with a fine, new sled.(5) It was a double-runner which his uncle, who owns a carriage factory in the city, had given him.(6) He named his new sled the Simoon(7) and almost had a fight with Tom Smith,(8) who said it was foolish to put such a name on a sled, but he kept on calling it the Simoon.(9)

At recess that day Frank invited the whole school to go for a coast, and the twelve boys and girls got on to the sled and away they went down the steep hill.(10) When recess was over, Miss Black, the teacher, rang the bell but not a scholar appeared. Thinking that the children had stopped to play on the way back from their slide, Miss Black went to the door and looked down the hill and rang the bell again. But not a scholar was in sight.(11) Then she was greatly astonished and began to be very angry,(12) for nothing like this had ever happened in all of her twenty-eight years as a teacher.(13) She waited and waited, but still no scholars appeared.(14) She stopped every team that came up the hill, but no one had seen anything of her school.(15)

She stayed at the schoolhouse and wondered what had become of her children until it was time to let out school(16) and then she went over to John Reed's, who lives nearest to the schoolhouse(17) and whose son and daughter were among the missing scholars.(18) Mr. Reed was greatly frightened at what Miss Black told him about the disappearance of her school(19) and immediately hitched up his horse to go in search of the lost children.(20) Just as he was driving out of the dooryard, the school appeared far down the hill.(21) It was almost dark before they got back to the schoolhouse.(22)

The idea which he considered expressed in each of these sections has been condensed in a short statement. These form a key for scoring. The statements corresponding to the sections in the portion of the test reproduced above are given below:

1. *Some children in the country attend school.*
2. *The schoolhouse is known as the Long Hill School.*
3. *It is situated on top of a long hill.*
4. *The pupils slide down hill once at recess in winter.*
5. *One day a boy brought to school a new sled.*
6. *His uncle had given it to him.*
7. *He named it the Simoon.*
8. *He almost had a fight with another boy.*
9. *This boy said the name was foolish.*
10. *At recess the pupils went for a slide.*
11. *At the end of recess no pupils appeared.*
12. *The teacher was astonished and angry.*
13. *Nothing like this had ever happened before.*
14. *After a long wait no scholars appeared.*
15. *No one in passing teams had seen her school.*
16. *She stayed at school until closing time.*
17. *Then she went to the nearest neighbor.*
18. *His children were among the scholars.*
19. *He was greatly frightened.*
20. *He started to search for the children.*
21. *Just then they appeared down the hill.*
22. *They reached the schoolhouse just before dark.*

For using this key he gives the following directions:¹⁹

¹⁹Brown's statement of these directions has been modified in order to make their meaning clear.

1. Each child's written reproduction should be carefully examined, and the number of points in the key which are reproduced by him should be determined and expressed as a percent of the total number in that portion of the selection read. For example, in the part read by a certain child, there may have been forty-eight points, and he may have reproduced twelve of these. The amount reproduced is, therefore, twenty-five percent of the amount read. This is called "quantity of reproduction". In arriving at a measure of quantity of comprehension, every idea reproduced by the child should be counted which, in most respects, is complete and which, in general, is correctly stated, even though some of the less important details are lacking. Credit for quantity of comprehension is given only when all elements of the idea expressed by the words in italics in the key are either expressed or plainly implied in the child's reproduction.

2. The reproductions should be examined a second time and only those ideas counted which are entirely correct in every respect and of which every detail is reproduced. This is called "quality of reproduction".

2. **The word-counting method.** In applying this method, a pupil's reproduction is examined and the words which do not correctly reproduce the selection read are crossed out. The pupil's score is the number of words remaining. The directions for crossing out words were essentially the same as those used by Starch in scoring his own silent reading tests. The scorers were directed to cross out the following classes of words:

- (a) Words which incompletely reproduce the thought.
- (b) Words which introduce new ideas.
- (c) Words which represent ideas reproduced elsewhere.
- (d) Superfluous connectives.

The scorers were, also, directed to bear constantly in mind that the aim of this method is to ascertain the number of words which actually reproduce the thought contained in the passage read. In order to secure independence on the part of the scorers when using the word-counting method, the lines of the reproductions were numbered. Sheets of ruled paper were then prepared with numbered lines. In scoring the reproductions, the words to be omitted in a line, when computing the pupil's score, were written on the corresponding line of the sheet of ruled paper. The number of words remaining in the line of the reproduction was then recorded in the right hand margin. The sum of these entries constituted a pupil's

score. No mark other than the numbers of the lines of the reproductions was made upon the pupil's test paper. Thus, the second scorer was not influenced in any way by the work of the first. The two independent scorings were reconciled by a third person, according to the rules given in the case of the idea-counting method, except that a difference of eight rather than of six was allowed before re-scoring was undertaken. This exception does not apply to the Memory Test.

Subjectivity of describing reproductions. An examination of the records of scoring the reproductions shows many differences of opinion on the part of the scorers. One scorer gave credit for certain words or ideas which the other scorer rejected, while the second scorer gave credit for words and ideas rejected by the first scorer. These differences of opinion tend to balance each other in the resulting scores but not entirely. For some reproductions, two persons will give the same score. For others, the two scores will differ. In a few cases the difference will be marked. Whenever there is a difference, at least one score, and probably both, involve an error.²⁰ Even when the two scores are identical both may involve an error.

Constant errors and variable errors. The scoring of reproductions even under favorable conditions, such as prevailed in this investigation, involves two types of errors—constant errors and variable errors. A constant error results in a scorer assigning scores which, in general, are too high or too low. A liberal attitude toward the reproductions will result in high scores. On the other hand, a conservative procedure will result in low scores. An indication of the presence of a constant error may be secured by comparing the averages of the two sets of scores assigned independently by two scorers to the same set of papers. Any differences in their general policy will be reflected by a difference between the averages of the two sets of scores. However, this difference cannot be considered to be an index of the magnitude of the constant error because both persons may be inclined to be liberal in their scoring, or both may be conservative, or one may be conservative and the other liberal.

Variable errors are indicated by the fact that in scoring one reproduction Scorer A will assign a score of 90, and Scorer B a score of 75; but in scoring a second reproduction Scorer A may assign a score of 60, and Scorer B a score of 80. This may happen although

²⁰A score is said to involve an error when it differs from the true score which is defined as the average of a large number of scores assigned by different persons.

Scorer B is, in general, more liberal than Scorer A. In studying the variable errors it is necessary to isolate them from the constant errors. Constant errors which affect the average of the scores assigned by either person do not affect the coefficient of correlation. Hence, it may be used as an index of the magnitude of the variable errors.

Tables I and II give data relative to both the constant and variable errors involved in the word-counting and in the idea-counting methods. Table I shows the facts for the first method and Table II for the second. The scorers are represented by letters. The numbers in the column headed "Difference of Average Scores" were obtained by subtracting the average of the scores assigned by the second scorer from the average of the scores assigned by the first scorer. A positive difference means that the first scorer gave, on the average, higher scores than the second. A negative difference has the opposite meaning. In some cases the difference closely approximates zero, but in others it is relatively large. This indicates that, for some scorers, the constant error is relatively large. One is justified in asserting that, on the basis of the possible constant error in the scores assigned to reproductions by a single scorer, no reliable inferences can be made concerning the differences in reading ability of two groups of pupils unless the differences between their average scores are large.

TABLE I. SUBJECTIVITY OF SCORING REPRODUCTIONS BY THE WORD-COUNTING METHOD

Test	Form	Grade	Number of scores	Scorers	Difference of average scores	r_{12}	P.E. Est. ₁₁	P.E. Est. ₁₁
								Average
Memory.....	I	IV	92	Y-C	-9.9	.89	4.5	.06
Memory.....	I	IV	27	Y-K	-5.1	.92	3.4	.04
Memory.....	II	IV	116	Y-C	-2.0	.90	3.3	.05
Memory.....	I	VII	123	Y-K	-7.5	.77	5.5	.05
Memory.....	II	VII	100	Y-C	-8.2	.84	3.9	.04
Memory.....	II	VII	31	Y-K	+4.1	.90	2.6	.03
Reproduction...	I	IV	94	L-K	+6.8	.97	3.1	.06
Reproduction...	II	IV	31	L-C	-1.6	.98	2.4	.06
Reproduction...	II	IV	68	L-K	+4.7	.92	4.2	.10
Reproduction...	I	VII	117	M-F	-0.5	.96	9.2	.06
Reproduction...	II	VII	113	F-C	-6.0	.97	5.5	.05
Brown.....	I	IV	111	T-M _y	+12.8	.83	6.1	.15
Brown.....	II	IV	110	T-M _y	+6.9	.88	4.8	.08
Starch (No. 7)...	I	VII	119	M-C	-5.8	.97	2.6	.07
Starch (No. 6)...	II	VII	121	M-C	-2.0	.97	2.1	.05

TABLE II. SUBJECTIVITY OF SCORING REPRODUCTIONS BY THE IDEA-COUNTING METHOD

Test	From	Grade	Number of scores	Scorers	Difference of average scores	r_{12}	P.E. Est. ₁₁	P.E. Est. ₁₁
								Average
Memory.....	I	IV	121	Y-P	+0.1	.95	1.1	.04
Memory.....	II	IV	116	Y-P	+0.6	.84	1.1	.05
Memory.....	I	VII	122	Y-P	+1.0	.89	1.6	.04
Memory.....	II	VII	128	Y-P	+0.6	.85	1.0	.04
Reproduction....	I	IV	94	F-P	-0.6	.94	1.6	.07
Reproduction....	II	IV	100	F-P	+0.7	.95	1.4	.08
Reproduction....	I	VII	116	F-P	-7.9	.91	5.6	.08
Reproduction....	II	VII	112	S-F	+0.7	.88	4.5	.10
Brown*.....	I	V	77	Cl-S	+0.4	.88	2.5	.10
Brown*.....	II	V	75	Cl-S	+1.5	.85	2.4	.11
Brown, Quantity	I	IV	112	P-C	+8.7	.69	8.4	.18
Brown, Quantity	II	IV	116	P-C	+7.8	.75	5.2	.16
Brown, Quality...	I	IV	113	P-C	-6.7	.68	6.1	.24
Brown, Quality...	II	IV	118	P-C	+0.1	.56	5.2	.30
Starch (No. 7)....	I	VII	122	S-Cl	-2.3	.92	1.6	.08
Starch (No. 6)....	II	VII	124	S-Cl	-1.0	.95	1.3	.08

*Brown I is *The Long Slide*; Brown II, *A Morning Adventure*.

It appears that a scorer is not always consistent with respect to his constant error. In Table I, Scorer Y and Scorer K show negative differences for two sets of papers and a positive difference for a third set. The same condition is exhibited by Scorer P and Scorer C in Table II. This reversal of policy may be due in part to differences in the character of the reproductions, but, doubtless, the instability of subjective judgment is also a factor.

In the column headed " r_{12} ", the coefficient of correlation between the two sets of scores is given. In the next column the probable error of estimate is given. This was calculated by the formula,²¹

$$P. E. Est_{11} = .6745 \sigma \sqrt{1-r_{12}^2}$$

²¹The probable error of estimate for two sets of related data is given by the formula $P. E. Est_{12} = .6745 \sigma_1 \sqrt{1-r_{12}^2}$ (See Yule, Introduction to the Theory of Statistics, Page 177.) In this formula r_{12} is the coefficient of correlation between the two sets of data and σ_1 is the standard deviation of the corresponding distribution. The probable error of estimate for the first set of scores (P. E. Est.₁) is a measure of the amount of change which would be necessary to bring these scores into perfect correlation with the other set of scores. Professor T. L. Kelley has shown that the correlation between one set of obtained scores and the corresponding true scores is given by the formula, $r_{1t} = \sqrt{r_{12}}$. Therefore, the formula, $P. E. Est_{11} = .6745 \sigma_1 \sqrt{1-r_{12}}$ gives the probable error of estimate of the first set of scores with respect to the corresponding set of true scores. A similar formula would give the probable error of estimate for the other set of scores. Since both sets of scores were assigned to the same set of reproductions, the best measure is the average of the two formulae. Hence, σ is the average of σ_1 and σ_2 .

As used here the probable error of estimate should be interpreted as a description of the magnitude of the variable errors or departures of the assigned scores from the corresponding true scores after the constant error has been eliminated. We may, therefore, speak of the probable error of estimate in this case as the *probable variable error of scoring*. A probable variable error of scoring of 3.4 means that, in general, the variable errors for the two scorers from whom the data were obtained are greater than 3.4 for fifty percent of the scores. It also means that for fifty percent of the scores the variable errors are less than 3.4.

The probable variable error of scoring cannot be given a definite significance except in comparison with the magnitude of the score with which it is to be associated. A probable error of 5 does not have the same meaning when associated with a score whose magnitude is 25 as it has when associated with a score of 100. It is, therefore, necessary to compare the probable variable error of scoring with the magnitude of the scores with which it is associated. The same degree of objectivity will result in larger variable errors of scoring for large scores than for small scores. Since the probable variable error of scoring which we have obtained is, itself, an "average" it may consistently be compared with the average score. This has been done in obtaining the quantities given in the last column of the table. The probable variable error of scoring has been divided by the average score. A quotient of .06 is to be interpreted as meaning that the chances are one to one that the score assigned to a paper will differ from the true score by as much as six percent of its magnitude.

In both tables, the coefficients of correlation are high in the sense that most of them differ only slightly from 1.00. With the exception of coefficients for "quality of reproduction" and "quantity of reproduction" of Brown's Silent Reading Test, only one is below .83. A number are above .90. There are four coefficients of .97. One is .98. With three exceptions, the number of cases on which these coefficients are based is sufficiently large so that the probable error of the coefficient of correlation due to sampling is relatively small. The description of the variable errors of scoring in terms of the probable variable error of scoring and the ratio of the probable variable error of scoring to the average suggest that these errors are much larger than might be concluded from a consideration of the coefficients of correlation. For example, in Table I the highest coefficient of correlation is .98 for the second form of the Experimental Repro-

duction Test in the fourth grade. The probable variable error of scoring is 2.4 units, which is six percent of the average score. This means that, in general, the chances are one to one that the score assigned to a pupil's reproduction in this group of papers will differ by at least six percent of its magnitude from the true score. This is the effect, only, of the variable error of scoring. The actual error of a pupil's score may be larger, due to the effect of the constant error on the part of the scorer.

It should also be noted that the highest coefficient of correlation is not always paired with the lowest ratio of the probable error of scoring to the average. In Table II, a ratio of .04 is obtained for three tests. The coefficients of correlation for these are .95, .89, and .85. In Table I, there are four ratios of .06. The corresponding coefficients of correlation are .89, .97, .98, and .96. The lowest ratio, .03, is associated with a coefficient of .90. Comparisons between the coefficients of correlation and the probable variable errors of scoring, likewise, show many cases of non-agreement. In Table I, the largest probable variable error, 9.2, corresponds to a coefficient of correlation of .96. The lowest coefficient of correlation, .77, corresponds to a probable variable error of 5.5. The smallest probable variable error, 2.1, corresponds to a coefficient of .97. This lack of agreement is due largely to differences in the magnitude of the scores.

The scoring of Brown's Silent Reading Test for quality and quantity of reproduction clearly involves the largest variable error. This is indicated both by the coefficient of correlation and by the probable variable error of scoring. If we exclude from our consideration these two scores of Brown's test, neither the idea-counting method nor the word-counting method is distinctly superior. In general, the word-counting method appears to involve a slightly smaller variable error when this error is considered in relation to the average score. However, both methods must be described as highly subjective. They involve a probable variable error of scoring of .06 or more in addition to a constant error which, in some cases, is probably large.

The scoring of Brown's test appears to be somewhat less objective than that of the others. This is especially true in the case of the word-counting method. In addition to the variable errors, this method appears to introduce a large constant error. The scores, "quantity of reproduction" and "quality of reproduction," which Brown recommends, are clearly less objective than the scores ob-

tained by either of the other methods. In fact, they are so highly subjective that their use cannot be defended.

Summary for describing reproductions. The description of reproductions involves large errors, both constant and variable. Even when the scoring is done under careful supervision reliable scores cannot be expected. For this reason, alone, silent reading tests requiring reproduction cannot be considered satisfactory. The method which Brown recommends for scoring reproductions appears to be inferior to both the word-counting method and the idea-counting method.

Scoring answers to questions. The scoring of the answers to the questions in the case of the Experimental Reproduction Tests and Fordyce's test is not perfectly objective unless an elaborate list of acceptable answers is prepared. This was done for both of these tests and, consequently, the scores used in this study may be considered objective in the sense that the scoring approximated uniformity. These tests, however, should not be considered as being perfectly objective when used independently by different persons who do not have access to elaborate directions for scoring.

Describing the quality of compositions. The scoring of the compositions for story value by means of the Willing Scale for Written Composition is not highly objective. Eighty-six compositions were scored independently by two persons. The difference between the averages of the two sets of scores was 6.7. The coefficient of correlation between the two sets of scores was .86. The probable variable error of scoring was 2.9 and the ratio of this to the average was .04. The magnitude of the variable error of scoring indicated by the probable error and its ratio to the average is less than that involved in either method of scoring the reproductions.

Time required for scoring test papers. All scorers kept a record of the time devoted to scoring the different tests. As we have indicated, care was exercised in the scoring and this probably tended to increase the time consumed. Furthermore, in the scoring of reproductions the procedure followed was not the most economical one. The average number of papers scored per hour is given in Table III. The most rapid scoring was done in the case of the questions of the Experimental Reproduction Tests. The scoring was nearly as rapid for Monroe's Standardized Silent Reading Tests and for the Pressey Test. The scoring of the tests requiring reproductions was relatively slow except in the case of Starch's Silent Reading Tests for ideas.

Average scores and standard deviations. In Tables IV and V, the average scores and standard deviations are given for each of the tests in each grade. The averages for the comprehension scores indicate that widely different units are used in describing the performances on the different tests. In the fourth grade the averages range from 6.2 for one method of scoring the Cross-Out Test to 87, the average index of comprehension yielded by the Curtis Silent Reading Test, No. 2. Even in the case of tests for which the unit is given the same name we have differences in magnitude. For example, the word is used as a unit in describing the reproductions. The average scores for tests requiring reproduction differ widely for the same pupils. In the seventh grade the average score for Form I of Starch's test is 40; for the Experimental Reproduction Test it is 155. The conditions under which these two tests are administered are not the same and this is, doubtless, one factor which causes the difference in the scores. Differences in the difficulty of the tests also tend to produce differences in the average scores. It is, however, likely that the units are not equivalent in the two cases. At least, they do not have equivalent interpretations when used as measures of comprehension.

TABLE III. AVERAGE NUMBER OF PAPERS SCORED
PER HOUR

Test	Method of Scoring	Grade	
		IV	VII
Monroe.....	Usual	48	53
Curtis.....	Usual	26	—
Brown.....	Word	15	—
Brown.....	Idea	24	—
Starch.....	Word	—	18
Starch.....	Idea	—	43
Reproduction.....	Word	15	11
Reproduction.....	Idea	26	8
Reproduction	Question	60	56
Cross-Out.....	Usual	21	39
Fordyce.....	Usual	27	28
Pressey.....	Usual	—	47
Vocabulary	Usual	20	16
Composition.....	Usual	—	13

TABLE IV. AVERAGE SCORES AND STANDARD DEVIATIONS FOR MEASURES OF COMPREHENSION

Test	Grade IV				Grade VII			
	Form I		Form II		Form I		Form II	
	Av.	σ	Av.	σ	Av.	σ	Av.	σ
Monroe.....	13.3	6.2	15.5	5.1	23.9	8.8	29.3	9.7
Courtis, Index.....	84.3	16.2	87.4	14.2	—	—	—	—
Courtis, Question.....	36.8	11.3	42.3	10.4	—	—	—	—
Courtis, Questions Correct	32.0	9.2	38.1	11.0	—	—	—	—
Brown, Quantity.....	47.7	20.9	29.8	14.5	—	—	—	—
Brown, Quality.....	30.4	17.6	17.9	11.7	—	—	—	—
Brown, Average.....	34.0	18.1	24.3	12.2	—	—	—	—
Brown, Efficiency.....	53.1	35.3	47.8	19.9	—	—	—	—
Brown, Words.....	42.4	23.4	57.4	21.5	—	—	—	—
Brown, Ideas.....	24.7	10.5	21.8	8.0	—	—	—	—
Starch, Words.....	—	—	—	—	40.9	18.3	38.1	22.3
Starch, Ideas.....	—	—	—	—	16.9	8.7	18.7	9.6
Reproduction, Question....	10.3	2.8	9.1	2.3	8.5	2.7	9.9	3.0
Reproduction, Ideas.....	20.8	11.1	17.8	10.4	65.3	33.5	42.5	23.6
Reproduction, Words.....	54.5	30.5	40.2	25.7	155.4	82.2	104.2	54.3
Cross-Out, C—W.....	6.2	5.0	8.5	5.9	16.9	7.1	18.1	8.2
Cross-Out, $\frac{C-W}{C+O}$	42.2	29.9	43.4	27.4	67.3	22.5	69.5	23.7
Fordyce.....	62.5	15.9	—	—	72.3	17.5	—	—
Pressey.....	—	—	—	—	13.9	3.3	14.1	3.2
Memory, Ideas.....	26.9	6.8	21.3	3.9	36.2	7.4	27.9	3.8
Memory, Words.....	77.7	19.6	76.0	13.4	104.3	19.8	91.0	13.1
Vocabulary.....	45.0	17.1	—	—	63.4	12.5	—	—
Composition.....	—	—	—	—	67.2	11.4	—	—

The non-equivalence of units is even more obvious in the case of the average rate scores. In four of the tests the pupil is engaged in continuous reading: Courtis Silent Reading Test, No. 2, Brown's Silent Reading Test, Starch's Silent Reading Tests, and the Experimental Reproduction Tests. The average rate scores for these tests exhibit differences sufficiently large to indicate that a word is not a constant unit for the measurement of the rate of reading. For example, the rate score for Form 3 of the Courtis Silent Reading Test is 153 words per minute. For Brown's Silent Reading Test the rate is 182 words per minute. Similar differences are to be found in the

TABLE V. AVERAGE SCORES AND STANDARD DEVIATIONS FOR MEASURES OF RATE

Test	Grade IV				Grade VII			
	Form I		Form II		Form I		Form II	
	Av.	σ	Av.	σ	Av.	σ	Av.	σ
Monroe.....	79.5	24.9	94.9	21.1	104.0	30.9	140.7	24.8
Courtis.....	150.0	47.9	153.1	55.2	—	—	—	—
Brown.....	164.6	60.3	182.5	78.5	—	—	—	—
Starch.....	—	—	—	—	193.0	56.5	202.8	88.9
Reproduction.....	151.6	77.7	154.7	77.1	218.3	82.7	216.3	70.2
Cross-Out.....	75.3	28.2	84.0	22.2	111.7	30.1	133.8	34.3
Fordyce, Words.....	125.0	27.0	—	—	179.0	35.5	—	—
Pressey.....	—	—	—	—	24.0	1.5	23.5	1.7
Composition (Number of words written.).....	—	—	—	—	218.6	85.5	—	—

seventh grade between Starch's Silent Reading Tests and the Experimental Reproduction Tests.

The rate scores for all of the tests are expressed in terms of words per minute. However, in the case of Monroe's Standardized Silent Reading Tests, The Cross-Out Silent Reading Tests, and the Pressey Silent Reading Test, the pupil does not do continuous reading. He must stop frequently to give responses. This, naturally, tends to reduce the rate scores. This is clearly shown in Table V. The rate scores for these tests are in most cases considerably less than rate scores in tests where the pupil does continuous reading. The difference is less marked in the seventh grade than in the fourth.

In Fordyce's Scale for Measuring Achievement in Reading, the pupil reads continuously, but the time allowance is such that a majority of the pupils complete the reading. Thus, they do not have an opportunity to give evidence of their rate of reading. This is the principal reason why the average rate scores for Fordyce's Tests are smaller than for the other tests in which the pupil does continuous reading.

The standard deviations also exhibit differences. Differences in the magnitude of the units would naturally affect the standard deviations as well as the averages. The standard deviation is also affected by the shape of the distribution. In a number of cases, the

distribution of scores does not approximate the normal shape. This is, doubtless, one factor affecting the differences between the standard deviations.

Equivalence of duplicate forms The facts given in Tables IV and V indicate that the forms of these tests are not equivalent. In some cases an effort was made to construct the different forms so that they would be equivalent. This is true of Monroe's Standardized Silent Reading Tests. A study²² planned to determine the degree of equivalence of these tests has indicated very definitely that they are not equivalent. The degree of non-equivalence revealed by that study is approximately that which is indicated here. The two forms of the Experimental Reproduction Tests, which were constructed without any preliminary study to determine their equivalence, appear to be as nearly equivalent as those of any other test in the list, as far as the rate is concerned. In the case of comprehension, there is considerable difference between the average scores. The two forms of the Cross-Out Tests were also constructed without much regard to equivalence and the average scores differ widely in most cases.

There is no published statement concerning the procedure followed by the authors of the other tests in order to secure equivalence of the duplicate forms. The average scores for the Courtis Silent Reading Test No. 2 do not differ widely. In fact, the two forms of this test appear to be the most nearly equivalent of any of the tests studied. The two Starch tests, No. 6 and No. 7, were not intended by the author to be equivalent forms. No. 7 (Form I) was intended to be more difficult, and lower average scores are, therefore, to be expected. This is what we find, except for the word-counting method of describing the reproductions. It is, however, obvious that it is difficult, or impossible, to construct duplicate forms which will be essentially equivalent, especially in the case of a small group of pupils. In addition to any lack of equivalence which may exist, the practise effect, due to one form being given after the other, would tend to produce differences between the average scores. The amount of this practise effect was not studied, since it was not pertinent to the major problem.

²²Monroe, Walter S. Report of Division of Educational Tests for 1919-20. University of Illinois Bulletin, Vol. XVIII, No. 21, Page 19.

Relation of vocabulary to difficulty. In an effort to determine whether the vocabulary of a selection tends to determine its difficulty, the selections read by pupils in tests requiring reproduction were analyzed. All the words occurring in each selection were listed and the frequency of each one determined. The number of words in each selection not occurring in Ayres' list of one thousand words was also determined. In the case of the selections which formed duplicate tests, the vocabularies were compared, and the number of words common to the two selections was found. The results of this study are given in Table VI. For the Courtis Silent Reading Test, No. 2, 16 percent of the vocabulary in Form 1 and 19 percent of the vocabulary in Form 2 are not found in the Ayres' list. The number of different words, or the vocabulary, of Form 1 is 37 percent of the length of the selection. This means that, on the average, each word in the selection is used nearly three times. In the case of Form 3, the number of different words is 44 percent of the total number of words in the selection. The number of words common to the two selections is 15 percent of the average number of words in the two selections. These facts show that for these two forms of the Courtis Silent Reading Test, No. 2, the two selections are approximately equivalent with respect to the percent of words not found in the Ayres' list. Form 3 contains a slightly larger percent of words not in this list. Such words will, in general, be unusual words unless they are proper names. Form 3 has a rela-

TABLE VI. ANALYSIS OF SELECTIONS READ BY PUPILS IN SILENT READING TESTS REQUIRING REPRODUCTION

Test	Words not in Ayres' list.	Different words.	Words common to both selections.
Courtis I.....	.16	.37	.15
Courtis III.....	.19	.44	—
Starch, No. 6.....	.30	.55	.10
Starch, No. 7.....	.31	.59	—
Brown			
Long Slide.....	.13	.37	.13
Morning Adventure.....	.14	.35	—
Old English			
Heroes I.....	.19	.43	.13
II.....	.19	.44	—
The Strike at Shane's I....	.19	.44	.13
II....	.19	.52	—

dively larger vocabulary, and makes a greater demand upon a pupil's acquaintance with words. The percent of words which are common to the two selections is surprisingly small in view of the simple character of the material and of the fact that the two selections are considered equivalent in difficulty.

A comparison of the facts contained in Table VI with those in Tables IV and V indicates that the explanation for the non-equivalence of the two forms of the same test is not to be found in the vocabularies of the two selections in the respective tests. Evidently, the difficulty of a selection is determined by some factor other than the actual words used.

Formation of composite scores. The scores yielded by the different tests are expressed in terms of different scales. Therefore, it is necessary to reduce them to a common scale before combining them to form composite scores. The procedure adopted was to choose as a base the scale of Monroe's Standardized Silent Reading Test I, Form 1, for the fourth grade and the scale of Test II, Form 1, for the seventh grade. All other scores were reduced to the scale of these tests. The formula for reducing the scores obtained from one scale to equivalent scores on another scale is as follows:

$$S_1 = \frac{\sigma_1}{\sigma_2} S_2 + (Av_1 - \frac{\sigma_1}{\sigma_2} Av_2)$$

In this formula, S_2 is the obtained score on Form 2 and S_1 is the equivalent score expressed in terms of the scale of Form 1. Av_1 refers to the average of the scores obtained from Form 1; Av_2 refers to the average of the scores obtained from Form 2. The standard deviation of the distribution of the Form 1 scores is σ_1 , and σ_2 is the standard deviation of the distribution of the Form 2 scores. This formula is based upon the usual assumption that corresponding deviations from averages are equal when expressed in terms of the standard deviation of the distribution; in other words, that

$$\frac{S_1 - Av_1}{\sigma_1} = \frac{S_2 - Av_2}{\sigma_2}$$

When this equation is solved for S_1 we obtain the formula as given above. The application of the above formula involves the determination of the numerical value of the ratio of $\frac{\sigma_1}{\sigma_2}$ by which the Form 2 score is to be multiplied and the determination of the numerical equivalent of the constant term of the formula (i. e., of the expression in parentheses). This latter numerical equivalent may be plus or

minus. When it is positive it is to be added and when negative it is to be subtracted.

After the scores were reduced to the same scale composite scores were formed by calculating the averages of certain groups of scores. Composite AI is the average of Monroe, Form 1 (comprehension), Curtis, Form 1 (answers correct), and Reproduction, Form 1 (answers correct). (In the seventh grade the Curtis Test was not given and this composite score includes only the other two tests.) Composite AII is obtained from the second form of these tests.²³ Composite BI is the average of Brown's Silent Reading Test (both quality and quantity scores), and the Experimental Reproduction Tests (ideas and words). In the seventh grade, Starch's Silent Reading Test²⁴ (ideas and words) is used in the place of Brown's test. Composite CI is the average of Composite AI and Composite BI. Composite BII and CII were obtained in a corresponding way from the second forms of these tests. Composite I is obtained by combining all Form 1 scores. Composite II is obtained by combining all Form 2 scores.

Reliability. Since, with the exception of Fordyce's Scale for the Measurement of Achievement in Reading, two forms of each test were given, it is possible to compute measures of the extent to which equivalent scores were yielded by the different forms of a test. It is also possible to compute the *probable error of measurement* which is a measure of the magnitude of the departures of the obtained scores from the corresponding true scores.²⁵ These departures are the variable errors of measurement. No account is taken of the constant error of measurement in the following discussion. In the case of the tests for which the scoring is subjective, the computed reliability is greater than the true reliability for the reason that the averages of two independent scorings were used instead of the scores assigned by one person.²⁶

Methods of determining reliability. In Tables VII and VIII, the reliability of these tests is described in terms of four quantities. (1) The coefficient of reliability is represented by the symbol, r_{12} , and is the coefficient of correlation between the two sets of scores yielded by the two forms of the test. (2) The index of reliability is represented by the symbol, r_1 . This quantity is the coefficient

²³In the case of the Curtis Tests, Form 3 was used instead of Form 2.

²⁴No. 7 is Form 1 and No. 6, Form 2.

²⁵A true score is defined as the average of the scores yielded by a large number of duplicate forms of a test.

²⁶See page 17 for the exact method used.

of correlation between one set of obtained scores and the set of corresponding true scores. The relation between the index of reliability and the coefficient of reliability is expressed by $r_{1t} = \sqrt{r_{12}}$. This formula was used in calculating the indices of reliability given in these two tables. (3) The probable error of measurement is represented by the symbol, P.E.M. This quantity was calculated by the formula, $P.E.M = .6745 \sigma \sqrt{1 - r_{12}}$.²⁷

The probable error of measurement (P. E.M) is a measure of the variable errors of measurement, or the differences between the obtained scores and the corresponding true scores. (4) The ratio of the probable error of measurement to the average of the scores from which it was calculated is represented by the symbol, $\frac{P.E.M}{Av.}$. Table VII gives information concerning the reliability of rate scores and Table VIII, the corresponding information for comprehension scores. In case the test was scored by more than one method, the information is given for all methods of scoring.

Probable error of r due to sampling. The coefficients of correlation, given in Tables VII and VIII and in the following tables, are subject to an error of sampling when interpreted with respect to the existence of relationship between the two sets of data from which they were derived. All of the correlations in the following tables are based on 80 cases in the fourth grade and 91 in the seventh

TABLE VII. MEASURES OF RELIABILITY, RATE

Test	Grade IV				Grade VII			
	r_{12}	r_{1t}	P.E.M	$\frac{P.E.M}{Av.}$	r_{12}	r_{1t}	P.E.M	$\frac{P.E.M}{Av.}$
Monroe I-II.....	.76	.87	11.3	.13	.63	.79	17.0	.11
Monroe I-III.....	.64	.80	13.6	.15	.55	.74	16.6	.11
Monroe II-III.....	.68	.82	11.8	.12	.69	.83	12.3	.09
Courtis.....	.85	.92	19.3	.13	—	—	—	—
Brown.....	.86	.93	26.0	.15	—	—	—	—
Starch.....	—	—	—	—	.62	.79	44.8	.23
Reproduction.....	.74	.86	39.5	.26	.45	.67	56.6	.26
Cross-Out.....	.68	.82	14.4	.18	.76	.87	15.8	.13
Pressey.....	—	—	—	—	.50	.71	1.1	.05

²⁷For explanation of this formula and the method of application see page 24.

TABLE VIII. MEASURES OF RELIABILITY, COMPREHENSION

Test	Grade IV				Grade VII			
	r_{12}	r_{11}	P.E.M	P.E.M	r_{12}	r_{11}	P.E.M	P.E.M
				Av.				Av.
Monroe I-II.....	.73	.85	3.0	.21	.69	.83	5.2	.20
Monroe I-III.....	.65	.81	3.4	.22	.60	.77	5.6	.20
Monroe II-III.....	.78	.88	2.5	.15	.61	.78	5.2	.17
Courtis, Index.....	.58	.76	9.9	.12	—	—	—	—
Courtis, No. of Questions.	.80	.89	4.6	.11	—	—	—	—
Courtis, No. of Questions Correct.....	.69	.83	5.7	.16	—	—	—	—
Brown, Quantity.....	.36	.60	14.2	.37	—	—	—	—
Brown, Quality.....	.19	.44	13.1	.54	—	—	—	—
Brown, Average.....	.32	.57	16.6	.55	—	—	—	—
Brown, Efficiency.....	.47	.69	20.1	.43	—	—	—	—
Brown, Words.....	.63	.79	13.7	.27	—	—	—	—
Brown, Ideas.....	.48	.69	6.6	.28	—	—	—	—
Starch, Words.....	—	—	—	—	.77	.88	9.7	.25
Starch, Ideas.....	—	—	—	—	.72	.85	4.8	.27
Reproduction, Questions..	.48	.69	1.9	.20	.60	.78	1.8	.20
Reproduction, Ideas.....	.54	.73	7.1	.37	.72	.85	15.1	.15
Reproduction, Words.....	.50	.71	19.5	.41	.87	.93	24.6	.19
Reproduction, Questions and Words.....	.44	.63	4.0	.30	.64	.80	4.6	.20
Cross-Out C-W.....	.62	.79	3.4	.47	.67	.82	4.6	.26
Cross-Out $\frac{C-W}{C+O}$52	.72	19.9	.47	.52	.72	14.7	.21
Pressey.....	—	—	—	—	.65	.81	1.9	.14
Memory, Ideas.....	.35	.59	4.3	.18	.56	.75	3.3	.10
Memory, Words.....	.40	.63	12.8	.17	.34	.58	13.3	.13

grade. In order to economize space, we give in Table IX probable errors due to sampling for various values of r . Most of the coefficients of correlation appearing in these tables are sufficiently large in comparison with the probable error due to sampling that they may be interpreted as indicating the existence of a distinct positive relationship. We are, however, more interested in securing a measure of the departure from perfect correlation. Hence, the probable error of measurement (P.E.M) is a much better index of the degree of reliability of a test than either r_{12} or r_{11} .

Reliability of the tests studied. Brown's Silent Reading Test, when scored in the way which he recommends, is the least reliable. The ratio of the probable error of measurement to the average is .54 for the quality and .55 for the average of quantity and quality.

TABLE IX. PROBABLE ERRORS OF THE COEFFICIENT OF CORRELATION (r_{12})
DUE TO USING A LIMITED NUMBER OF CASES*

r_{12}	P. E.
.1	.0798
.2	.0774
.3	.0734
.4	.0677
.5	.0605
.6	.0516
.65	.0466
.70	.0411
.75	.0353
.80	.0290
.85	.0224
.90	.0153
.95	.0079

*80 in Grade IV and 91 in Grade VII.

The "efficiency score" has a ratio of .43. The scoring of this test by means of either the idea-counting method or the word-counting method results in scores that are more reliable. Considering both rate and comprehension, the most reliable test is the Curtis Silent Reading Test, No. 2. For rate, the index of reliability is .92 and the ratio of the probable error of measurement to the average is .13. Three comprehension scores are used in connection with this test. The number of questions answered is shown to be the most reliable.

The probable error of measurement and the ratio of the probable error of measurement to the average score indicate a degree of reliability for the rate scores yielded by Monroe's Standardized Silent Reading Tests which is surprisingly high, considering the character of the tests. This is particularly true in the seventh grade. With the exception of Pressey's Silent Reading Test, they are the most reliable. In the fourth grade, the reliability is exceeded only by Curtis's Silent Reading Test, No. 2. In Monroe's Standardized Silent Reading Tests a pupil does not read continuously but is forced to stop at the end of each exercise and answer a question. According to the rules for scoring these tests, a pupil receives no credit for an exercise unless he has completed his reading of it to the extent of recording his answer. The increments added to a rate score for doing additional exercises are relatively large, particularly in Test II. Thus, a pupil who has failed only in recording his answer to an exercise receives a score which does not indicate his rate of reading. His score is the same as that of the pupil who has just barely completed the preceding exercise. In all of the other tests, with the exception of Pressey's Silent Reading Test, the pupil's rate score represents the actual amount read. In view of these facts,

it is surprising to find that Monroe's Standardized Silent Reading Tests yield rate scores which have such a high degree of reliability. The figures which are given may be affected somewhat by the fact that these tests proved too short and a considerable number of pupils made perfect scores.

In general, the degree of reliability is higher in the seventh grade than in the fourth. Exact comparisons cannot be made because identical tests were not given in the two grades; but where similar tests were given the results for the seventh grade show a distinctly higher degree of reliability. This may be due to a superiority in the tests for the seventh grade or it may be due to the fact that the increased maturity of the pupils causes them to be less variable in their performances.

The degree of unreliability shown in Tables VII and VIII is distressingly high. As we have indicated, the ratio of the probable error of measurement to the average probably furnishes the most significant statement of the degree of unreliability. Brown's Test, scored by any method, appears to be so highly unreliable that it should be rejected. In interpreting the figures in Table VIII it should be borne in mind that the actual degree of unreliability is somewhat larger than that indicated because the element of subjectivity in scoring has been largely eliminated. It appears that individual scores yielded by these tests are very imperfect measures of reading ability. However, the variable errors involved do not affect, to the same degree, the scores of classes or larger groups. Although the scores yielded by these tests must be considered as having only a very limited significance in the case of individual pupils, they are much more significant for groups of pupils.

Both the Experimental Reproduction Tests and the Cross-Out Tests were merely experimental. The reproduction tests were intentionally so. It was desired to ascertain whether a crude reproduction test, such as might be constructed by a teacher and administered directly from a supplementary reader, would yield results as reliable as tests more carefully constructed and more conveniently arranged. These tests are shown to be among the least reliable, with the exception of Brown's Silent Reading Test. This is to be expected; but the difference in reliability, particularly in the seventh grade, is not marked. In fact, the Experimental Reproduction Tests exhibit a relatively high degree of reliability in the measurement of comprehension. Thus, the reliability of a crude test of

this type is only slightly less than that of tests whose construction was more refined.

Discrimination. The distributions of the rate scores yielded by the different tests indicate that certain tests fail to yield scores which discriminate between a number of pupils with respect to rate of reading. Form 3 of Monroe's Standardized Silent Reading Tests I and II is clearly too short. In the seventh grade 58 percent of the pupils and in the fourth grade 27 percent completed the test. All such pupils received the maximum rate score. The distributions for Forms 1 and 2 of this test contain no such extreme deviations from the normal shape, although Form 2 of Test I and Form 1 of Test II cannot be said to approximate closely the normal distribution.

The Cross-Out Tests yield distributions which exhibit many irregularities and which cannot be said to do more than suggest the normal distribution. As was to be expected, a large percent of the pupils completed the reading of the selection in the case of the Fordyce test. Forty-nine percent of the pupils in the fourth grade and 29 percent in the seventh grade received the maximum rate score. The Pressey Test proved too short for the time allowed. Seventy-six percent completed Form 1 and 56 percent completed Form 2. The Curtis, Brown, Starch, and Experimental Reproduction Tests yielded rate scores which formed distributions closely approximating the normal shape. A few irregularities were exhibited by the Experimental Reproduction Tests and by Brown's test.

As judged by the shape of the distribution of the rate scores, the Curtis Silent Reading Test, No. 2, exhibits the least lack of discrimination. The Cross-Out, Pressey, Fordyce, and Form 3 of Monroe's tests exhibit such great departures from the normal distribution that they must, obviously, fail to discriminate properly with respect to the rate of reading for a considerable number of pupils.

In the case of comprehension, the distributions of scores for Monroe's Standardized Silent Reading Tests closely approximate the normal. The third form appears to have been a little too easy; but, in other respects, the irregularities exhibited by the distributions cannot be considered to indicate a serious lack of discrimination. The index of comprehension for the Curtis Silent Reading Test, No. 2, fails to discriminate properly between a number of pupils. Both the number of questions answered and the number of questions answered correctly approach more nearly the normal distribution.

TABLE X. CORRELATIONS WITH TEACHER RATING

Test	Rate		Comprehension	
	Grade IV	Grade VII	Grade IV	Grade VII
Monroe I.....	.38	.25	.60	.32
Monroe II.....	.34	.35	.64	.50
Monroe III.....	.43	.26	.63	.39
Courtis I Index.....			.29	
Courtis I Questions.....			.29	
Courtis I Questions Correct.....			.41	
Courtis I Words per minute.....	.55			
Courtis III Index.....			.45	
Courtis III Questions.....			.38	
Courtis III Questions Correct.....			.51	
Courtis III Words per minute.....	.51			
Brown I Quantity.....			.30	
Brown I Quality.....			.29	
Brown I Average.....			.34	
Brown I Efficiency.....			.58	
Brown I Words.....			.64	
Brown I Ideas.....			.55	
Brown I Words per minute.....	.53			
Brown II Quantity.....			.01	
Brown II Quality.....			.05	
Brown II Average.....			.05	
Brown II Efficiency.....			.36	
Brown II Words.....			.48	
Brown II Ideas.....			.45	
Brown II Words per minute.....	.42			
Starch I Words.....				.46
Starch I Ideas.....		.19		.46
Starch I Words per minute.....				
Starch II Words.....				.51
Starch II Ideas.....				.49
Starch II Words per minute.....		.41		
Reproduction I Questions.....			.42	.34
Reproduction I Ideas.....			.31	.47
Reproduction I Words.....			.35	.23
Reproduction I Words per minute.....	.36	.26		
Reproduction II Questions.....			.36	.51
Reproduction II Ideas.....			.51	.49
Reproduction II Words.....			.43	.46
Reproduction II Words per minute.....	.32	.16		
Cross-Out I C-W.....			.39	.21
Cross-Out I $\frac{C-W}{C+O}$39	.27
Cross-Out I Words per minute.....	.36	.40		
Cross-Out II C-W.....			.56	.34
Cross-Out II $\frac{C-W}{C+O}$46	.20
Cross-Out II Words per minute.....	.43	.36		
Fordyce.....	.49	.13	.24	.39
Pressey I.....		.29		.46
Pressey II.....		.08		.37
Composite AI.....			.58	.40
Composite AII.....			.60	.55
Composite BI.....			.44	.53
Composite BII.....			.34	.58
Composite CI.....			.59	.51
Composite CII.....			.56	.63
Composite I.....	.55	.45		.58
Composite II.....	.51	.36		.58

This is particularly true of the latter. The distributions for the Brown, Starch, and Experimental Reproduction Tests exhibit many irregularities; but there is in all cases a distinct resemblance to the normal distribution. A few of the distributions approach very closely the normal one. Others contain rather marked departures from it. In the case of Brown's test, the distributions for the quality scores exhibit greater departures than the distributions for the quantity scores.

Comparison with teachers' ratings. All scores, both rate and comprehension, were correlated with the ratings in silent reading given by the teacher. The coefficients of correlation were calculated, also, for certain composite scores. These coefficients of correlation are given in Table X. With the exception of one coefficient for the second form of Brown's test, all coefficients are positive and in general sufficiently large to indicate a distinct positive relationship between the test scores and the teachers' ratings. Rate of reading correlates more highly with the teachers' rating in the fourth grade than in the seventh. For rate, the average of the coefficients, not including the composite scores, is 43 in the fourth grade and 26 in the seventh. The average of the coefficients for comprehension, not including the composite scores, is 40 in the fourth grade and 44 in the seventh.

In the fourth grade, comprehension, as measured by Monroe's Standardized Silent Reading Tests, correlates most highly with the teachers' ratings. In fact, the coefficients for the three forms of this test equal or exceed all of those for the composite scores. In the seventh grade this test does not exhibit as high correlations with teachers' ratings. Neither do its rate scores correlate as highly with teachers' ratings as the rate scores yielded by some other tests. It is interesting to note that the correlation between the second form of Brown's Test for "quantity of reproduction" and "quality of reproduction" is essentially zero. For Form 1 the correlations for these two scores are lower than the correlations for any other scores. This suggests that Brown's method for scoring his test is undesirable. The correlations of the composite scores with teachers' ratings indicate that, in the fourth grade, teachers judge silent reading ability more on the basis of the pupils' ability to answer questions than of their ability to reproduce. In the seventh grade, the teachers give greater weight to the pupils' ability to reproduce or to tell what has been read.

Correlation of comprehension with memory. In those tests which require the pupil to answer questions from memory or to

TABLE XI. CORRELATION OF COMPREHENSION WITH MEMORY

Test	Grade IV				Grade VII			
	Ideas		Words		Ideas		Words	
	I	II	I	II	I	II	I	II
Brown I Quantity.....	.32	.21	.39	.27	—	—	—	—
Brown II Quantity.....	.27	.20	.23	.20	—	—	—	—
Brown I Quality.....	.36	.16	.36	.16	—	—	—	—
Brown II Quality.....	.19	.11	.14	.08	—	—	—	—
Starch I Ideas.....	—	—	—	—	.36	.27	.32	.31
Starch II Ideas.....	—	—	—	—	.39	.24	.34	.25
Starch I Words.....	—	—	—	—	.37	.28	.33	.47
Starch II Words.....	—	—	—	—	.24	.33	.45	.34
Reproduction I Questions.....	.46	.35	.46	.33	.43	.34	.33	.26
Reproduction II Questions.....	.27	.38	.30	.36	.51	.27	.44	.20
Reproduction I Ideas.....	.42	.41	.46	.41	.53	.43	.55	.36
Reproduction II Ideas.....	.39	.24	.35	.28	.56	.40	.50	.35
Reproduction I Words.....	.37	.41	.35	.41	.45	.35	.53	.33
Reproduction II Words.....	.41	.28	.38	.27	.54	.40	.53	.39
Monroe I.....	.31	.33	.28	.27	.49	.47	.34	.35
Monroe II.....	.25	.33	.24	.25	.52	.48	.45	.24
Monroe III.....	.31	.35	.35	.38	.37	.33	.40	.26
Maximum.....	.46	.41	.46	.41	.56	.48	.55	.47
Minimum.....	.19	.11	.14	.08	.24	.24	.32	.20
Average.....	.34	.29	.33	.28	.44	.36	.42	.32

TABLE XII. CORRECTED COEFFICIENTS OF CORRELATION OF COMPREHENSION WITH MEMORY

Test	Grade IV		Grade VII	
	Ideas	Words	Ideas	Words
Brown Quantity.....	.67	.66	—	—
Brown Quality.....	.68	.54	—	—
Starch Ideas.....	—	—	.51	.66
Starch Words.....	—	—	.39	.90
Reproduction Questions.....	.75	.72	.72	.75
Reproduction Ideas.....	.97	.86	.77	.86
Reproduction Words.....	.98	.88	.62	.77
Monroe I-II.....	.57	.47	.80	.82
Monroe I-III.....	.68	.61	.72	.83
Monroe II-III.....	.61	.58	.71	.75

reproduce the passage read, it would seem that a pupil's ability to remember would materially affect his comprehension score. In order to ascertain the extent to which ability to remember does affect the comprehension score yielded by such tests, the pupils were given the memory test²⁸ described on page 7. In this test a selection was read to the pupils and they were asked to reproduce the story from memory. The coefficients of correlation between the memory scores and the comprehension scores for silent reading tests are given in Table XI. It is significant that none of these coefficients are large. The first three tests listed in this table require the pupil to give his performances from memory. Monroe's Standardized Silent Reading Tests do not appear to make any considerable demand upon the pupil's memory; he has the passage before him and can read it and re-read it if he desires. If any memory is involved it is immediate in character. It is significant that the coefficients of correlation for this test closely approximate those for other tests.

Corrected coefficients of correlation. The measures yielded by these tests involve variable errors. It has been shown in our consideration of the reliability of these tests that these errors are relatively large for the reproduction tests. The presence of these variable errors tends to reduce the coefficients of correlation, and it is possible that the coefficients of correlation given in Table XI do not represent the true relation between comprehension and memory.

When two forms of both tests have been given to the same pupils it is possible to compute a corrected coefficient of correlation which is free from the effect of the variable errors of measurement. This has been done by means of the following formula:²⁹

$$r_{pq} = \frac{\sqrt{(r_{p_1q_2})(r_{p_2q_1})}}{\sqrt{(r_{p_1p_2})(r_{q_1q_2})}}$$

r_{pq} here indicates the true correlation between two series of measures, p and q , of the facts A and B.

p_1 and p_2 are two independent measures of A.

q_1 and q_2 are two independent measures of B.

$r_{p_1q_2}$ is the correlation obtained from the first measure of A and the second measure of B.

$r_{p_2q_1}$ is the correlation obtained from the second measure of A and the first measure of B.

²⁸It is assumed that this test measures ability to remember.

²⁹Thorndike, E. L. "An Introduction to Mental and Social Measurements." New York. Teachers College, Columbia University, 1916. Page 179.

$r_{p_1p_2}$ is the correlation between the two measures of A.

$r_{q_1q_2}$ is the correlation between the two measures of B.

In applying this formula the factors of the numerator are obtained from Table XI. For example, in calculating the corrected coefficient of correlation for Brown's Silent Reading Test with memory, $r_{p_1q_2}$ is the coefficient of correlation of Brown I with Memory II. This is given as .21. The coefficient of correlation of Brown II with Memory I, is $r_{p_2q_1}$. This is given as .27. The factors of the denominator are the reliability coefficients of the two tests. These are to be found in Table VIII. They are .36 for Brown's Silent Reading Tests and .35 for the Memory Tests. Substituting these values in the formula,

$$\begin{aligned} r_{pq} &= \frac{\sqrt{.21 \times .27}}{\sqrt{.36 \times .35}} \\ &= \sqrt{.45} \\ &= .67 \end{aligned}$$

This is the first entry of the first column of Table XII.

A study of the corrected coefficients given in Table XII indicates that, in the case of the Experimental Reproduction Tests in the fourth grade, the correlation between Memory and the scores based upon the pupil's reproduction is very high. For ideas it is .97. For words it is .88. For Brown's Silent Reading Tests the correlation is not as high. In fact, it closely approximates that for Monroe's Standardized Silent Reading Tests. In the seventh grade the correlation of Memory with Monroe's Standardized Silent Reading Tests is higher than that for either Starch or the Experimental Reproduction Tests, although the difference is not marked in the case of the latter. It, therefore, appears that in the seventh grade memory is not a major factor in determining the comprehension scores of tests which require reproduction unless it is also the determining factor in the case of tests which do not appear to involve memory. The statement which has been made with reference to reproduction tests, that they measure the ability to read *and remember*, does not appear to be justified by the facts which are presented here.

Correlation of comprehension with vocabulary. In Table XIII, we give the coefficients of correlation between the comprehension scores and the scores obtained from the vocabulary test. In the fourth grade most of the coefficients are negative, but all of them cluster closely around zero. This means that, measured by the tests used,

TABLE XIII. COEFFICIENTS OF CORRELATION BETWEEN VOCABULARY AND COMPREHENSION

Test	Grade IV	Grade VII
Monroe I.....	.02	.22
Monroe II.....	-.03	.22
Monroe III.....	-.02	.13
Courtis I Index.....	-.20	—
Courtis I No. of Questions.....	.19	—
Courtis I Questions Correct.....	.10	—
Courtis III Index.....	-.20	—
Courtis III No. of Questions.....	.06	—
Courtis III Questions Correct.....	.15	—
Starch I Words.....	—	.31
Starch I Ideas.....	—	.32
Starch II Words.....	—	.29
Starch II Ideas.....	—	.22
Brown I Quantity.....	-.11	—
Brown I Quality.....	-.12	—
Brown I Average.....	.14	—
Brown I Words.....	.01	—
Brown I Ideas.....	-.04	—
Brown II Quantity.....	-.23	—
Brown II Quality.....	-.21	—
Brown II Average.....	-.16	—
Brown II Words.....	-.15	—
Brown II Ideas.....	-.19	—
Reproduction I Questions.....	-.15	.14
Reproduction I Ideas.....	-.10	.17
Reproduction I Words.....	-.09	.13
Reproduction I Questions.....	.12	.19
Reproduction I Ideas.....	-.04	.24
Reproduction I Words.....	-.04	.26
Cross-Out I C-W.....	-.07	.18
Cross-Out I $\frac{C-W}{C+O}$	-.05	.08
Cross-Out II C-W.....	-.09	.16
Cross-Out II $\frac{C-W}{C+O}$	-.02	.01
Fordyce.....	.04	.13
Pressey I.....	—	.21
Pressey II.....	—	.00
Composite AI.....	-.02	.23
Composite AII.....	.01	.20
Composite BI.....	-.08	.32
Composite BII.....	-.20	.28
Composite CI.....	-.05	.26
Composite CII.....	-.13	.25
Composite I.....	—	.30
Composite II.....	—	.21

there is no relation between a pupil's vocabulary and his ability to read. It is, of course, obvious that, in order to read, a pupil must be acquainted with words. It is, therefore, impossible to believe that vocabulary is not a factor in the reading process. The facts presented here probably mean that, in the fourth grade, vocabulary is not a determining factor and the pupil's ability to read depends primarily upon abilities other than the extent of his acquaintance with words. In the seventh grade the coefficients are all positive but none of them are large. This probably means that, in the seventh grade, vocabulary is a minor factor in determining the pupil's comprehension. It is, of course, possible that the vocabulary test used does not measure the extent of a pupil's acquaintance with words.

Correlation of cancellation scores with measures of rate of reading.

In Table XIV, the coefficients of correlation for the scores yielded by the Cancellation Test with measures of rate of silent reading are given. With few exceptions, these coefficients are positive but small. In general, they are slightly smaller in the seventh grade than in the fourth grade. In most cases, there does not seem to be any marked relationship between ability to do the Cancellation Test and the rate of silent reading. One might expect a distinct positive relationship between the Cross-Out Silent Reading Tests and the Pressey Silent Reading Tests. It does, however, appear that the relationship which exists with respect to these tests is greater than that which exists for Monroe's Silent Reading Tests.

The table also includes coefficients of correlation for the scores yielded by the Cancellation Test with the comprehension scores yielded by the Cross-Out Tests. The coefficients are, likewise, small, two of them being slightly negative. It appears, therefore, that the ability to strike out letters from words is not related to the ability called for by the Cross-Out Tests.

Correlation of comprehension with written composition.

Another measure of a pupil's vocabulary is secured from his written composition. The pupils in the seventh grade were asked to write a composition on an exciting experience. (See page 10.) In Table XV, we give the coefficients of correlation between measures of comprehension and two measures of these written compositions, the number of words written and the story value. The number of words which a pupil writes in such an exercise is, undoubtedly, an index of his writing vocabulary. It is, of course, possible that his writing

TABLE XIV. CORRELATION OF CANCELLATION SCORES WITH MEASURES OF RATE OF READING AND WITH THE CROSS-OUT TESTS

Test	Grade IV		Grade VII	
	Cancellation		Cancellation	
	I*	II*	I	II
Monroe I.....	.28	.14	.07	.20
Monroe II.....	.26	.15	.10	.18
Monroe III.....	.23	.20	-.03	.22
Courtis I.....	.12	.15		
Courtis III.....	.20	.14		
Brown I.....	.25	.16		
Brown II.....	.07	.08		
Starch I.....			-.06	-.01
Starch II.....			-.15	.03
Reproduction I.....	.22	.13	.05	.25
Reproduction II.....	.13	.08	.10	.22
Cross-Out I.....	.20	.15	.18	.25
Cross-Out II.....	.23	.13	.14	.33
Fordyce, No. of Words.....	.30	.12	-.08	.11
Pressey I.....			-.03	.11
Pressey II.....			.06	.15
Cross-Out I C-W.....	.08	.01	.21	.18
Cross-Out I C-W C+O.....	.02	-.03	.11	.05
Cross-Out II C-W.....	.17	.06	.16	.17
Cross-Out II C-W C+O.....	.14	.03	.11	-.01

*In Cancellation Test I, the words containing both "a" and "t" were marked; in Test II, those containing both "e" and "r."

vocabulary and his reading vocabulary are not closely related. The coefficients of correlation, in Table XV, show that there is little or no relation existing between measures of comprehension and the number of words which were written in these compositions. Even in the case of comprehension scores based upon the number of words and the number of ideas contained in reproductions, the coefficients of correlation fail to indicate the existence of any marked relationship. In fact, the coefficients of correlation for measures of comprehension gained through reproduction are lower, in most cases, than the coefficients of correlation of the number of words written with the comprehension scores derived from Monroe's Standardized Silent Reading Tests.

A higher degree of correlation is indicated between the "story value" and the measures of comprehension. Some of the coefficients of correlation are sufficiently large to indicate a distinct positive relationship between these two traits. It is not unlikely that this relationship can be explained in terms of a common general factor, such as general intelligence.

Inter-correlation between tests. Since in each grade all of the tests were given to the same pupils, it is possible to calculate the coefficients of correlation between scores yielded by the different tests. These are given in the appendix. The magnitude of the coefficients of correlation is influenced by the reliability of the scores and, therefore, does not truthfully reflect the relationship which exists between the scores yielded by the different tests. In order to secure more accurate indices of the relationship existing between traits measured by the different tests, the corrected coefficients of correlation have been calculated by means of the formula given on page 41. Since the factors of both numerator and denominator of the formula are square roots, it is impossible to calculate corrected coefficients when one of the raw coefficients is negative. This

.TABLE XV. CORRELATION OF COMPREHENSION WITH WRITTEN COMPOSITION, SEVENTH GRADE, 90 PUPILS

Test	Number of words written	Story value
Monroe I.....	.18	.29
Monroe II.....	.23	.33
Monroe III.....	.24	.31
Starch I Words.....	.10	.31
Starch I Ideas.....	.07	.28
Starch II Words.....	.14	.36
Starch II Ideas.....	.09	.33
Reproduction I Questions.....	.12	.24
Reproduction I Ideas.....	.22	.14
Reproduction I Words.....	.22	.18
Reproduction II Questions.....	-.07	.11
Reproduction II Ideas.....	.26	.37
Reproduction II Words.....	.28	.43
Cross-Out I C-W.....	.13	.23
Cross-Out I $\frac{C-W}{C+O}$09	.11
Cross-Out II C-W.....	.16	.06
Cross-Out II $\frac{C-W}{C+O}$04	.11
Fordyce Percent.....	.12	.12
Pressey I.....	.10	.29
Pressey II.....	.05	.18

accounts for the fact that certain corrected coefficients are not given in Tables XVI and XVII. It will be noted in these tables that, occasionally, a coefficient greater than 1.00 is given. This is due to chance errors in the raw coefficients of correlation which, in turn, are due to the fact that a sample of the total population was used in calculating them. The corrected coefficients are, in general, larger than the corresponding raw coefficients.

Table XVI gives the corrected coefficients for the comprehension scores. A significant characteristic of this table is the variation in the degree of intercorrelation between the tests. For example, Monroe's Standardized Silent Reading Test I correlates very highly with the number of questions answered correctly on the Curtis Silent Reading Test, No. 2. It correlates less highly with the other two scores of this test. The degree of its correlation with the other tests is moderately low. It is significant that the corrected coefficients of correlation between the two tests requiring reproduction are not higher. For example, the highest coefficient of correlation between Brown's test and the Experimental Reproduction Test I is .79. The lowest is .26. The corrected coefficient of correlation between the scores obtained by the word-counting method is .33; for the idea-counting method the coefficient of correlation is .62. The highest correlation between Brown's test and the Experimental Reproduction Test I is for the number of questions answered correctly. In the seventh grade, the corrected coefficients of correlation between the question scores yielded by the Experimental Reproduction Test II and Starch's Silent Reading Test are as high as those obtained from the reproductions. Both Starch's test and the Experimental Reproduction Test correlate nearly as highly with Monroe's Standardized Silent Reading Test as with each other. A number of the coefficients of correlation for the Cross-Out Test are relatively high. It correlates most highly with Monroe's Standardized Silent Reading Test. In general, the coefficients are higher for the scores obtained by C—W than for $\frac{C-W}{C+O}$. The former is probably the better plan of scoring.

Table XVI appears to bear out the usual assumption that different silent reading tests measure different phases of silent reading ability. It is very obvious, in a number of cases, that the same traits are not measured by different tests. However, it should be noted that these differences exist for tests that are similar in structure as well as for tests which possess marked differences in struc-

TABLE XVI. CORRECTED COEFFICIENTS OF CORRELATION OF MEASURES OF COMPREHENSION OF READING

TESTS	Grade IV			Grade VII			Grade IV Courts			Grade IV Brown				Grade VII Starch		Grade IV Reprod.			Grade VII Reprod.			Grade IV Cross-Out		Grade VII Cross-Out		Grade VII Pressy			
	I-II	III-I	III-II	I-II	I-III	II-III	Index of Comp.	No. of Questions	Questions Correct	Quantity	Quality	Average	Words	Ideas	Words	Ideas	Questions	Ideas	Words	Questions	Ideas	Words	Questions	Ideas	Words		Questions	Words	Ideas
Monroe I-II.....							.57	.66	.92			.65	.59	.72	.72	.44	.65	.49	.39	.69	.44	.87	.64	.60					
I-III.....							.62	.72	.97	.00		.73	.67	.68	.59	.63	.59	.58	.48	.79	.33	.84	.72	.66					
II-III.....							.72	.57	.89			.73	.70	.75	.74	.66	.48	.48	.62	.52	.46	.75	.52	.91	.65				
Courts										.40																			
Index of comprehension	.57	.62	.72							.55	.38	.50	.49																
No. of Questions	.66	.72	.57							.15		.57	.56																
Questions Correct.....	.92	.97	.89							.18		.79	.73																
Brown																													
Quantity.....			.00				.40																						
Quality.....							.40		.18																				
Average.....							.55		.15																				
Words.....	.65	.73	.73				.38																						
Ideas.....	.59	.67	.70				.50		.73																				
Starch																													
Words.....				.72	.68	.75																							
Ideas.....				.72	.59	.74																							
Reproduction																													
Questions.....	.55	.63	.66				.59		.76	.44		.61	.43	.79	.72	.57	.51												
Ideas.....	.47	.59	.48				.30		.47	.48		.39	.55	.69	.62	.49	.58												
Words.....	.44	.56	.48				.38		.35	.52	.59	.26	.47	.33	.68	.54	.51												
Cross-Out																													
C-W.....	.69	.79	.75				.63		.66	.28		.22	.19	.81	.71	.56	.63	.67	.27	.47	.58	.53	.46						
C-W.....	.44	.33	.52				.55		.41	.57	.61	.55	.71	.71	.52	.50	.52	.33	.42	.65	.48	.44							
C+O.....																													
Pressy.....				.69	.66	.83																							
Composite																													
A.....	.97	.97	.97				.53		.66	.79	.80	.84	.88	.48	.80	.91	.80	.63	.60	.53	.91	.51	.48						
B.....	.60	.42	.44				.42		.68	.38		.68	.67	.86	.86	.80	.78	.77	.80	.85	.87	.88	.78						
C.....	.80	.86	.82				.52		.62	.80	.43	.88	.88	.48	.80	.91	.80	.89	.83	.84	.82	.72							
General.....	.84	.84	.84				.84		.84	.84	.43	.88	.88	.48	.80	.91	.80	.89	.83	.84	.82	.72							

TABLE XVII. CORRECTED COEFFICIENTS OF CORRELATION OF MEASURES OF RATE OF READING

	Grade IV Monroe I- II III	Grade VII Monroe I- II III	Grade IV Courtis	Grade IV Brown	Grade VII Starch	Grade IV Reprod. Reprod.	Grade VII Reprod.	Grade IV Cross Out	Grade VII Cross Out	Grade VII Pressey	Grade IV Comp.	Grade VII Comp.	Grade VII Comp. A
Monroe I-II.....			.75	.69	.66	.74	.83	.91	.72	.81	.90	.94	.73
I-III.....			.82	.73	.64	.84	.87	.94	.61	.85	.94	.95	.77
II-III.....			.72	.73	.62	.66	.62	.95	.66	.84	.90	.91	.65
Courtis.....	.75	.82	.72	.84		.90		.76			.94		
Brown.....	.69	.73	.73	.84		.91		.68			.89		
Starch.....		.66	.64	.62			.72		.57	.57		.85	.86
Reproduction.....	.74	.84	.66	.91	.72			.73	.67	.41	.90	.95	1.04
Cross-Out.....	.91	.94	.95	.68	.57	.73	.67			.47	.83	.83	.64
Pressey.....		.81	.85	.84	.57		.41		.47			.57	.60
Composite.....	.90	.94	.90	.89	.85	.90	.95	.83	.83	.57			.93
Composite A.....		.73	.77	.65	.86		1.04		.64	.60		.93	

ture. In fact, the variations in these corrected coefficients of correlation are so erratic that one is inclined to be skeptical of any conclusions which may be drawn from them with reference to the functions of the different tests.

The corrected coefficients for the rate scores are given in Table XVII. These are, in general, higher than those for comprehension. In general, the correlation between tests in which the pupil reads continuously is higher than between one test in which the pupil reads continuously and another in which his reading is not continuous. However, the correlation between Monroe's Standardized Silent Reading Test I and the Cross-Out Test, in the fourth grade, is as high as that for any of the other tests. The fact that some of the tests were too short and failed to discriminate between a considerable number of pupils probably accounts for the fact that a number of coefficients of correlation are not higher. An examination of this table indicates that the rate score secured by means of Monroe's Standardized Silent Reading Tests is a true measure of the pupil's rate of reading.

Correlation of single tests with composites. In Tables XVI and XVII, the corrected coefficients of correlation for each test with certain composite scores are given. These, in general, are larger than the coefficients of correlation between single tests. In the fourth grade, composite A for comprehension is the average of Monroe, comprehension, Curtis, answers correct, and Reproduction, answers to questions. In the seventh grade, the Curtis test was not given and this composite includes only the other two tests. Composite B for comprehension is the average of the comprehension scores derived from reproductions. In the case of Brown's Silent Reading Tests, both quality and quantity are used. In the other cases, the scores obtained by both the idea-counting method and the word-counting method are used. Composite C is the average of composite A and composite B. The general composite is formed by combining all of the scores obtained.

Monroe's Standardized Silent Reading Tests are shown to correlate very highly with composite A. The correlation with composite B is very much less, as might be expected. The rate scores derived from this test also correlate very highly with the general composite scores. In fact, with the exception of Pressey's test, the correlation of single tests with the composite scores is very high. It appears, therefore, that each of the tests yields rate scores which

may be accepted as correlating very highly with the true rate of silent reading. The scores derived from the Experimental Reproduction Tests in the fourth grade correlate more highly with composite B than those derived from Brown's Silent Reading Test. In the seventh grade, the correlations between Starch's test and composite B are slightly higher than those for the Experimental Reproduction Tests. It appears, however, that the Experimental Reproduction Tests yield approximately as valid measurements of ability to comprehend as are secured by means of the other tests which, presumably, have been devised with greater care.

SUMMARY OF CONCLUSIONS.

1. The scoring of reproductions is so highly subjective that a silent reading test requiring reproduction of material read cannot be considered satisfactory.

2. Brown's Silent Reading Test is very unreliable for both comprehension and rate. This is true, even when the average of two independent scores is used as a measure of comprehension.

3. The correlation between scores yielded by the memory test and comprehension scores based upon reproductions is only slightly higher than that existing between the scores derived from the memory test and the comprehension scores yielded by Monroe's Standardized Silent Reading Test. This makes doubtful the usual assumption that measures of comprehension based upon reproductions are affected by the pupil's ability to remember.

4. Correlation between extent of vocabulary and ability to read is surprisingly low. There is little, if any, relation between these two abilities.

5. The intercorrelations between tests indicate that different tests measure slightly different traits; but it is surprising to find, in a few instances, a high degree of correlation existing between scores yielded by tests which exhibit marked differences in structure.

6. There appears to be a higher degree of correlation between the story value of written compositions and comprehension than between the number of words written and the measures of comprehension. This is true even when the measures of comprehension are based upon reproductions and the reproductions are described in terms of the number of words or number of ideas reproduced.

7. In the measurement of rate of silent reading, the Curtis Silent Reading Test No. 2, is shown to have the highest degree of reliability. Monroe's Standardized Silent Reading Tests, which

were intended to yield only very crude measures of rate of silent reading, are shown to be among the most reliable tests.

8. In measuring comprehension, the Courtis Silent Reading Test, No. 2, is the most reliable.

9. The coefficient of reliability is shown not to be a satisfactory measure of reliability.

10. Comparisons with teachers' ratings indicate that, in the fourth grade, teachers tend to judge silent reading ability on the basis of the pupil's ability to answer questions. In the seventh grade, teachers give greater weight to the pupil's ability to reproduce or tell what they have read.

Correlation with composites. In Tables XVI and XVII, the corrected coefficients of correlation of each test with the composite scores are given. These, in general, are larger than the correlations between single tests. Monroe's Standardized Silent Reading Test correlates very highly with composite A. This means that this test, which is very simple to administer, yields measures of essentially the same traits as are secured by means of this composite, which in the fourth grade involves three scores and in the seventh, two scores. The correlation with composite C and with the general composite is also high. In fact, with the partial exception of Starch's Test, no other correlations are as high as these two composites of the Monroe Silent Reading Tests. It, therefore, appears, as judged by composite scores, that this test yields measures of comprehension which agree more closely with the composite measures secured from this group of tests than any other single test. The correlations for rate are also high.

THE UNIVERSITY OF ILLINOIS

THE STATE UNIVERSITY

URBANA

DAVID KINLEY, Ph.D., LL.D., *President*

The University Includes the Following Departments

THE GRADUATE SCHOOL

THE COLLEGE OF LIBERAL ARTS AND SCIENCES (Ancient and Modern Languages and Literatures; History, Economics, Political Science, Sociology, Philosophy, Psychology, Education; Mathematics; Astronomy; Geology; Physics; Chemistry; Botany, Zoology, Entomology; Physiology, Art and Design)

THE COLLEGE OF COMMERCE AND BUSINESS ADMINISTRATION (General Business, Banking, Insurance, Accountancy, Railway Administration, Foreign Commerce; Courses for Commercial Teachers and Commercial and Civic Secretaries)

THE COLLEGE OF ENGINEERING (Architecture; Architectural, Ceramic, Civil, Electrical, Mechanical, Mining, Municipal and Sanitary, Railway Engineering, and General Engineering Physics)

THE COLLEGE OF AGRICULTURE (Agronomy; Animal Husbandry; Dairy Husbandry; Horticulture and Landscape Gardening; Agricultural Extension; Teachers' Course; Home Economics)

THE COLLEGE OF LAW (Three-year and four-year curriculums based on two years and one year of college work respectively)

THE COLLEGE OF EDUCATION

THE CURRICULUM IN JOURNALISM

THE CURRICULUMS IN CHEMISTRY AND CHEMICAL ENGINEERING

THE SCHOOL OF RAILWAY ENGINEERING AND ADMINISTRATION

THE SCHOOL OF MUSIC (four-year curriculum)

THE LIBRARY SCHOOL (two-year curriculum for college graduates)

THE COLLEGE OF MEDICINE (in Chicago)

THE COLLEGE OF DENTISTRY (in Chicago)

THE SCHOOL OF PHARMACY (in Chicago; Ph.G. and Ph.C. curriculums)

THE SUMMER SESSION (eight weeks)

EXPERIMENT STATIONS AND SCIENTIFIC BUREAUS: U. S. Agricultural Experiment Station; Engineering and Experiment Station; State Laboratory of Natural History; State Entomologist's Office; Biological Experiment Station on Illinois River; State Water Survey; State Geological Survey; U. S. Bureau of Mines Experiment Station.

The library collections contain May 1, 1922, 523,230 volumes and 120,131 pamphlets. For catalogs and information address

THE REGISTRAR

Urbana, Illinois



BULLETINS OF THE BUREAU OF EDUCATIONAL RE-
SEARCH, COLLEGE OF EDUCATION, UNIVERSITY
OF ILLINOIS, URBANA, ILLINOIS.

	<i>Price.</i>
No. 1. Buckingham, B. R. Bureau of Educational Research, Announcement, 1918-1915
No. 2. First Annual Report25
No. 3. Bamesberger, Velda C. Standard Requirements for Memorizing Literary Material50
No. 4. Holley, Charles E. Mental Tests for School Use. (Out of print)50
No. 5. Monroe, Walter S. Report of Division of Educational Tests for 1919-2025
No. 6. Monroe, Walter S. The Illinois Examination.....	.50
No. 7. Monroe, Walter S. Types of Learning Required of Pupils in the Seventh and Eighth Grades and in the High School15
No. 8. Monroe, Walter S. A Critical Study of Certain Silent Reading Tests.....	.50
No. 9. Monroe, Walter S. Written Examinations and Their Improvement. (In preparation).....	.50

LIBRARY OF CONGRESS

0 021 728 714 5

Hollinger Corp.
pH 8.5